Multilevel Coding Schemes for Compute-and-Forward with Flexible Decoding

Brett Hern and Krishna Narayanan

Department of Electrical and Computer Engineering

Texas A&M University

College Station

TX 77843, U.S.A

Abstract

We consider the design of coding schemes for the wireless two-way relaying channel when there is no channel state information at the transmitter. In the spirit of the compute and forward paradigm, we present a multilevel coding scheme that permits computation (or, decoding) of a class of functions at the relay. The function to be computed (or, decoded) is then chosen depending on the channel realization. We define such a class of functions which can be decoded at the relay using the proposed coding scheme and derive rates that are universally achievable over a set of channel gains when this class of functions is used at the relay. We develop our framework with general modulation formats in mind, but numerical results are presented for the case where each node transmits using the QPSK constellation. Numerical results with QPSK show that the flexibility afforded by our proposed scheme results in substantially higher rates than those achievable by always using a fixed function or by adapting the function at the relay but coding over GF(4).

Index Terms

Network coding, multilevel coding, two-way relaying, compute-and-forward

I. INTRODUCTION

Physical layer network coding (PLNC) or Compute and Forward is a new paradigm in wireless networks where each relay in a network decodes a function of the transmitted messages and broadcasts the value of this function to the other nodes in the network. This has been shown to provide significant increase in achievable rates for some networking problems [1], [2], [3]. For a recent and approachable tutorial/survey of the key ideas behind PLNC with reliable decoding, we refer readers to [4]. For another broad tutorial/survey of PLNC results with an eye to practical implementation, we refer readers to [5].
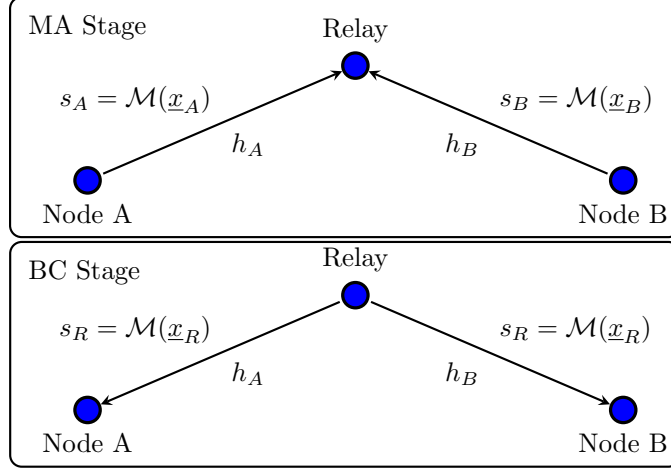
Fig. 1. System model showing a of two-way relay channel with PLNC.

An example of such a problem where compute and forward has been shown to be effective is the two-way relaying system shown in Fig. 1. Here, node A has data to send to node B and vice versa. The relay R is included to assist in this communication, and it is assumed that there is no direct link between nodes A and B. Near optimal coding schemes have been designed to maximize the exchange rate for the case where there is no fading in the channel in [2], [3], [6]. Building on results from [7], these authors derive an upper bound on the capacity of $\frac{1}{2}\log(1 + snr)$ and show that with lattice coding and lattice decoding a rate of $\frac{1}{2}\log(\frac{1}{2} + snr)$ is achievable. This problem has also been studied for case where there is fading in the channel, but each node perfectly knows the fading coefficients for each network link in [8]. It has been shown that near-optimal performance can be obtained at high signal-to-noise ratio (SNR) if each transmitter inverts its channel prior to transmission. The authors in [9] apply lattices with list decoding to the two way relaying problem with a direct link between nodes A and B. Finally, compute and forward schemes for multiple input multiple output channels have been considered in [10].

In this paper, the complex channel coefficients h_A and h_B are assumed to be perfectly estimated at each receiver but unknown to each transmitter. For this scenario, the authors in [11] introduce a scheme called denoise-and-forward which uses channel dependent denoising functions at the relay to minimize the symbol error probability. The relay chooses denoising functions so that the distance profile for constellation points with different labels is optimized. This improves the symbol error rate for transmissions between nodes A and B, however, denoising is performed purely at the symbol level. There is no natural extension to include error correction at the relay.

Recently, a scheme called compute-and-forward, which allows both adaptation of decoding functions and error correction at the relay has been presented in [12]. In this scheme, the relay decodes an integer combination of the transmitted codewords, where the integer combination is adapted according to the channel gains. They show that such a scheme can be implemented using nested lattice codes to take advantage of the duality between modulus arithmetic in prime order fields and the modular operations of lattice decoding. Their scheme requires the construction of infinite

dimensional lattice codes which is not practical. The results in [12] are extended in a remarkable way in [13], where an algebraic framework is provided to design lattices over principal ideal domains. However, their proposed coding scheme is also based on large dimensional lattice codes.

In this paper, we propose a compute and forward scheme based on multilevel coding (MLC). Unlike the coding schemes in [12], [13], our proposed scheme does not result in a lattice code and uses only linear codes over small prime fields (for example, binary linear codes), and can therefore be implemented with lower encoding and decoding complexity. Yet, it facilitates error correction for a larger class of decoding functions than those proposed in [12]. This is because the class of functions for our scheme is derived from the large set of non-singular square matrices over $\mathbb{F}_p$ in place of the set of non-zero elements in large prime order fields. To the best of our knowledge, such an idea of using multilevel coding and exploiting the linearity over the prime field to adaptively decode linear functions of transmitted codewords is new. Another important contribution in this paper is that our proof for the achievability of rates with the proposed multilevel coding scheme requires a non-trivial extension of the proof of achievability of rates for multilevel coding for the point to point case.

This paper is organized as follows. The key elements of the problem are outlined in Section II. Our proposed solution is detailed in Section III. An achievable rate for the proposed scheme during the MA stage is given in Section IV. These rates are numerically determined for an example where nodes A and B transmit using a QPSK constellation in Section V. Simulation results for a regular LDPC code are shown to corroborate the information theoretic results. Key results are reiterated in Section VI.

Throughout this paper we will use the following naming conventions. Vectors or sequences will be denoted by underlined variables such as $\underline{x}$. Random variables will be denoted by upper case variables such as X, while their outcomes will be represented by lowercase variables. Matrices will be represented by capital boldface letters such as $\mathbf{X}$. Subsets will be denoted by capital scripted letters such as $\mathcal{X}$. If a variable is associated with a specific node, this will be indicated by a subscripted capital letter like x_A. MLC sometimes requires us to split a data sequence into subsequences for parallel encoding and transmission over separate bit levels. Variables associated with a specific bit level will be indicated by a superscript like x^k. A specific element of a vector or sequence will be referred to by an index in brackets like $\underline{x}[n]$.

II. PROBLEM DESCRIPTION

Each node in the relay network is assumed to be half-duplex, so communication is split into two stages, a multiple access (MA) stage and a broadcast (BC) stage. We assume perfect synchronization between the transmitters and mainly focus on the MA stage in this paper.

A. Multiple Access Stage

Nodes A and B each encode their binary messages $\underline{u}_A$ and $\underline{u}_B$ into codewords $\underline{v}_A \in \mathcal{C}_A$ and $\underline{v}_B \in \mathcal{C}_B$ where $\mathcal{C}_A$ and $\mathcal{C}_B$ are the codebooks used at the nodes A and B respectively. These codewords are mapped to sequences of

symbols $\underline{s}_A, \underline{s}_B \in \mathcal{Q}^N$ with $|\mathcal{Q}| = 2^\ell$. The relay receives noisy observations of the sum of these symbol sequences according to

$$\underline{y}_R = h_A \underline{s}_A + h_B \underline{s}_B + \underline{w}_R \qquad (1)$$

where h_A and h_B are complex fading coefficients, and $\underline{w}_R$ is complex additive white Gaussian noise (AWGN). This induces an effective constellation $\mathcal{Q}_R$ at the relay defined by

$$\mathcal{Q}_R = \{q_R \in \mathbb{C} | q_R = h_A q_A + h_B q_B, \ q_A, q_B \in \mathcal{Q}\}. \qquad (2)$$

B. Adaptive Decoding at the Relay

The main idea proposed in this paper is the construction of a coding scheme such that the relay can reliably decode some function of $\underline{v}_A$ and $\underline{v}_B$ for a desired set of channel conditions $\mathcal{H} \subset \mathbb{C}^2$. Specifically, we jointly design codes $\mathcal{C}_A$ and $\mathcal{C}_B$ and a set of decoding functions $\mathcal{F}$ such that, for any $(h_A, h_B) \in \mathcal{H}$, there exists $f \in \mathcal{F}$ such that the relay can reliably decode $f(\underline{v}_A, \underline{v}_B)$ from $\underline{y}_R$. We require that node A (B) must be able to unambiguously decode $\underline{v}_B$ ($\underline{v}_A$) from the output of $f(\underline{v}_A, \underline{v}_B)$ with its knowledge of $\underline{v}_A$ ($\underline{v}_B$). For a given $f \in \mathcal{F}$, we will define an induced codebook at the relay as the codebook corresponding to f i.e.

$$\mathcal{C}_{f,R} = \{f(\underline{v}_A, \underline{v}_B) | \underline{v}_A \in \mathcal{C}_A, \ \underline{v}_B \in \mathcal{C}_B\}. \qquad (3)$$

It is important to understand the structure of $\mathcal{C}_{f,R}$ since the probability of error in decoding $f(\underline{v}_A, \underline{v}_B)$ from $\underline{y}_R$ depends on h_A, h_B, and $\mathcal{C}_{f,R}$. The main advantage of our proposed scheme is that it guarantees that choosing one codebook $\mathcal{C}_A$ and $\mathcal{C}_B$ at the transmitter can result in a good induced codebook $\mathcal{C}_{f,R}$ for a class of functions $\mathcal{F}$. More specifically, it guarantees $\mathcal{C}_{f,R}$ is a member of the ensemble of random coset codes which is an optimal ensemble for achieving the uniform input information rate for the equivalent channel between $f(\underline{v}_A, \underline{v}_B)$ and $\underline{y}_R$ for all $f \in \mathcal{F}$. We restrict our attention to classes of functions $\mathcal{F}$ which are applied componentwise at the relay.

The broadcast stage is fairly standard and is identical to that considered in [2], [3].

III. PROPOSED SCHEME

A. Multilevel Encoder

The system model for the multilevel encoder for nodes A and B and the channel model for the MA stage is shown in Fig. 2. The encoder at nodes A and B uses MLC with a different coset of the *same linear* code $\mathcal{C}$ used at each bit level. For a detailed description of MLC and achievable rates for the point to point channel see [14].

The encoder is described as it pertains to node A to simplify notation. First, the message $\underline{u}_A$ is split into sub-vectors $\underline{u}_A^1, ..., \underline{u}_A^\ell$ which form rows of an $\ell \times K$ matrix

$$\mathbf{U}_A = \begin{bmatrix} \underline{u}_A^1 \\ \vdots \\ \underline{u}_A^\ell \end{bmatrix}. \qquad (4)$$
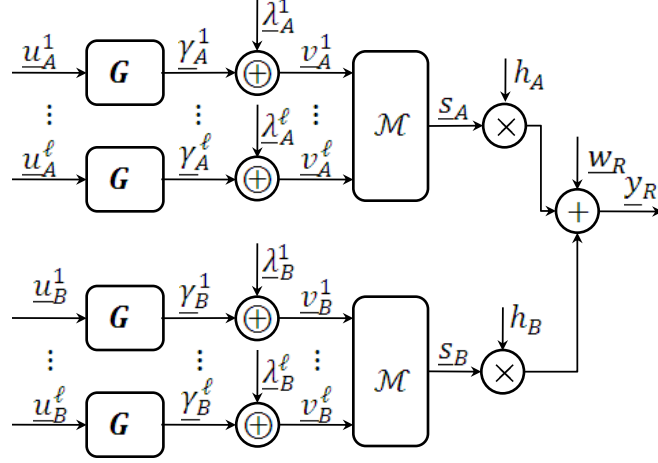
Fig. 2. Block Diagram of MLC Coset Encoders for MA Stage.

Each $\underline{u}_A^k$, $\{1, ..., \ell\}$ is encoded with a linear code $\mathcal{C}$ with generator matrix $\mathbf{G}$ to get codewords $\underline{\gamma}_A^1, ..., \underline{\gamma}_A^\ell$. These codewords from the rows of an $\ell \times N$ matrix

$$\mathbf{\Gamma}_A = \mathbf{U}_A \mathbf{G} = \begin{bmatrix} \underline{\gamma}_A^1 \\ \vdots \\ \underline{\gamma}_A^\ell \end{bmatrix}. \tag{5}$$

Finally, a random binary vector $\underline{\lambda}_A^k$ is added to each $\underline{\gamma}_A^k$. Each $\underline{\lambda}_A^k$ can be thought of as coset leaders of a random coset of the original linear code. We obtain a codeword of a random coset given by $\underline{v}_A^k = \underline{\lambda}_A^k \oplus \underline{\gamma}_A^k$, $k \in \{1, .., \ell\}$. The random coset leaders form an $\ell \times N$ matrix

$$\mathbf{\Lambda}_A = \begin{bmatrix} \underline{\lambda}_A^1 \\ \vdots \\ \underline{\lambda}_A^\ell \end{bmatrix}. \tag{6}$$

The resulting coset codewords $\underline{v}_A^k$ form the rows of a binary $\ell \times N$ matrix $\mathbf{X}_A$ given by

$$\mathbf{X}_A = \mathbf{U}_A \mathbf{G} \oplus \mathbf{\Lambda}_A = \begin{bmatrix} \underline{v}_A^1 \\ \vdots \\ \underline{v}_A^\ell \end{bmatrix} = [\underline{x}_A[1], ..., \underline{x}_A[N]]. \tag{7}$$

Thus each code $\mathcal{C}_A^k$, $k \in \{1, ..., \ell\}$ will be a different coset of $\mathcal{C}$. The k_{th} row $\underline{v}_A^k$ of $\mathbf{X}_A$ is then a codeword of $\mathcal{C}_A^k$. We use the two variables $\underline{x}_A[n]$ and $\underline{v}_A^k$ to refer to the n_{th} column and k_{th} row of $\mathbf{X}_A$ respectively because it will simplify our notation later. It should be mentioned here that much of the intuition about the main result in the paper is best obtained by ignoring the fact that cosets are used at each layer and simply considering the use of identical linear codes at each level in the MLC scheme. The coset matrix $\mathbf{\Lambda}_A$ is included to symmetrize the effective channel at the relay (i.e. $\mathbf{\Lambda}_A$ is necessary for the proofs to be correct).
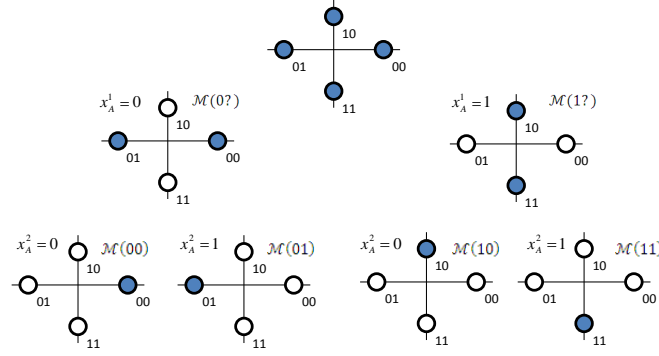
Fig. 3.   MLC address mapping example using QPSK.

The n_{th} binary address vector $\underline{x}_A[n] \in \mathbb{F}_2^\ell$ maps to a symbol $\underline{s}_A[n] \in \mathcal{Q}$ through the use of a symbol mapping function $\mathcal{M} : \mathbb{F}_2^\ell \to \mathcal{Q}$. An example of such a mapping function is given in Fig. 3 where $\mathcal{Q}$ is the QPSK constellation. As shown, the mapping function is usually derived by partitioning the set of signaling points in $\mathcal{Q}$ into equal sized subsets. Let $\mathcal{S} \subseteq \{1, ..., \ell\}$ be the subset of elements of $\underline{x}_A$ which are fixed. Then we define the output of $\mathcal{M}$ with these input bits as a subset of points from $\mathcal{Q}$ according to

$$\mathcal{M}(\{x_A^k | k \in \mathcal{S}\}) = \{q \in \mathcal{Q} | q = \mathcal{M}(\{x_A^k | k \in \mathcal{S}\}, \{b^i | i \in \overline{\mathcal{S}}\}), \ b^i \in \{0, 1\}\}. \tag{8}$$

This means that the returned subset of constellation points is the subset whose address vectors are equal to the known bits for all indexes, $\mathcal{S}$. The output of $\mathcal{M}(\{x_A^k | k \in \mathcal{S}\})$ is $2^{\ell - |\mathcal{S}|}$ constellation points.

In this example of Fig. 3 if $\mathcal{S} = \{2\} \subseteq \{1, 2\}$ and $x_A^2 = 1$, then

$$\mathcal{M}(\{x_A^k | k \in \mathcal{S}\}) = \mathcal{M}(?1) = \{\mathcal{M}(01), \mathcal{M}(11)\} = \{-1, -j\}.$$

Here, $|\mathcal{S}| = 1$ and $\ell = 2$. Therefore $\mathcal{M}$ returns $2^{2-1} = 2$ constellation points.

B. Adaptive Decoding at the Relay

As mentioned previously, the goal of the proposed scheme is to allow the relay to decode a function of the transmitted codewords. Similar to the compute and forward scheme, our scheme utilizes the linearity of the base code $\mathcal{C}$ and the fact that the relay knows Λ_A and Λ_B. If nodes A and B encode their messages as described, the set of decoding functions $\mathcal{F}$ which the relay can use for decoding is defined as follows.

Define $\mathcal{D}$ as the set of $\ell \times \ell$ binary matrices which are invertible over $\mathbb{F}_2$. The set of functions we consider is given by

$$\mathcal{F} = \{f : \mathbb{F}_2^\ell \times \mathbb{F}_2^\ell \to \mathbb{F}_2^\ell | f(\underline{x}_A, \underline{x}_B) = [\mathbf{D}_A \mathbf{D}_B] \begin{bmatrix} \underline{x}_A \\ \underline{x}_B \end{bmatrix}, \ \mathbf{D}_A, \mathbf{D}_B \in \mathcal{D}\}. \tag{9}$$

Therefore a given $f \in \mathcal{F}$ is defined by some $\mathbf{D}_A, \mathbf{D}_B \in \mathcal{D}$ from which the relay should attempt to decode a matrix $\mathbf{X}_{f,R}$ given by

$$\mathbf{X}_{f,R} = [\mathbf{D}_A \mathbf{D}_B] \begin{bmatrix} \mathbf{X}_A \\ \mathbf{X}_B \end{bmatrix}. \tag{10}$$

Due to the linearity of $[\mathbf{D}_A, \mathbf{D}_B]$ and $\mathbf{G}$, we can express the desired matrix $\mathbf{X}_{f,R}$ as

$$\mathbf{X}_{f,R} = [\mathbf{D}_A \mathbf{D}_B] \begin{bmatrix} \mathbf{X}_A \\ \mathbf{X}_B \end{bmatrix} = [\mathbf{D}_A \mathbf{D}_B] \begin{bmatrix} \mathbf{U}_A \mathbf{G} \oplus \mathbf{\Lambda}_A \\ \mathbf{U}_B \mathbf{G} \oplus \mathbf{\Lambda}_B \end{bmatrix}$$

$$= [\mathbf{D}_A \mathbf{D}_B] \begin{bmatrix} \mathbf{U}_A \mathbf{G} \\ \mathbf{U}_B \mathbf{G} \end{bmatrix} \oplus [\mathbf{D}_A \mathbf{D}_B] \begin{bmatrix} \mathbf{\Lambda}_A \\ \mathbf{\Lambda}_B \end{bmatrix}$$

$$= [\mathbf{D}_A \mathbf{D}_B] \begin{bmatrix} \mathbf{U}_A \\ \mathbf{U}_B \end{bmatrix} \mathbf{G} \oplus [\mathbf{D}_A \mathbf{D}_B] \begin{bmatrix} \mathbf{\Lambda}_A \\ \mathbf{\Lambda}_B \end{bmatrix}$$

$$= \mathbf{U}_{f,R} \mathbf{G} \oplus \mathbf{\Lambda}_{f,R}. \tag{11}$$

Here, we see that the matrix $\mathbf{X}_{f,R}$ can be written in terms of an effective message $\mathbf{U}_{f,R}$ and coset matrix $\mathbf{\Lambda}_{f,R}$ which can be computed separately based on f. Thus the rows of $\mathbf{X}_{f,R}$ are codewords from a different coset code of $\mathcal{C}$. Note that f is applied elementwise to the sequences $\underline{s}_A$ and $\underline{s}_B$.

For clarification, consider the case of $\ell = 2$. Let a function f_1 be defined by $\mathbf{D}_A = \mathbf{D}_B = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$. Writing the vectors $\underline{x}_A$ and $\underline{x}_B$ as $[x_A^1 \; x_A^2]^T$ and $[x_B^1 \; x_B^2]^T$ respectively, we see that $\underline{x}_{f_1,R} = [x_{f_1,R}^1 \; x_{f_1,R}^2]^T$ is given by

$$\underline{x}_{f_1,R} = f_1(\underline{x}_A, \underline{x}_B) = \begin{bmatrix} 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \end{bmatrix} \begin{bmatrix} x_A^1 \\ x_A^2 \\ x_B^1 \\ x_B^2 \end{bmatrix}.$$

This corresponds to the binary XOR function given by $f_1(\underline{x}_A, \underline{x}_B) = [x_A^1 \oplus x_B^1, x_A^2 \oplus x_B^2]^T$.

Define another function f_2 using $\mathbf{D}_A = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$ and $\mathbf{D}_B = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$. This is the rotated-XOR function given by $f_2(\underline{x}_A, \underline{x}_B) = [x_A^1 \oplus x_B^2, x_A^2 \oplus x_B^1]^T$.

Recall from (7), that $\underline{v}^k[n] \Leftrightarrow x^k[n]$. Thus using f_1 at the relay corresponds to decoding $[\underline{v}_A^1 \oplus \underline{v}_B^1]$ and $[\underline{v}_A^2 \oplus \underline{v}_B^2]$. Similarly, applying f_2 at the relay corresponds to decoding $[\underline{v}_A^1 \oplus \underline{v}_B^2]$ and $[\underline{v}_A^2 \oplus \underline{v}_B^1]$.

To illustrate the importance of choosing the decoding function f depending on (h_A, h_B), consider an example with $\mathcal{Q} = \{1, j, -1, -j\} = \{\mathcal{M}(00), \mathcal{M}(01), \mathcal{M}(11), \mathcal{M}(10)\}$ (i.e. QPSK with Gray Labeling). Further, let $h_A = e^{j\theta_A}$ and $h_B = e^{j\theta_B}$, and let $\theta = \theta_A - \theta_B$ be the phase difference. Consider the decoding functions

$$f_1(\underline{x}_A, \underline{x}_B) = [x_A^1 \oplus x_B^1, x_A^2 \oplus x_B^2]$$

$$f_2(\underline{x}_A, \underline{x}_B) = [x_A^1 \oplus x_B^2, x_A^2 \oplus x_B^1].$$
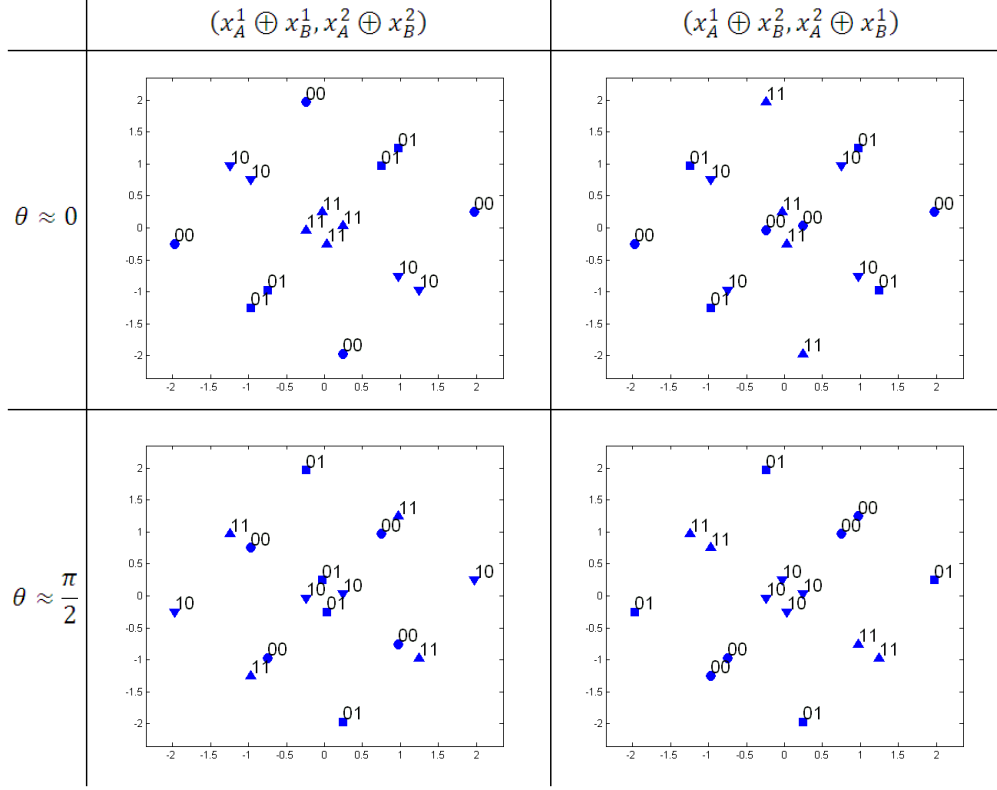
Fig. 4. Effective constellation at relay for different values of θ.

The resulting constellation $\mathcal{Q}_R$ at the relay is shown for different values of θ in Fig. 4. Note that the complex coordinates of the constellation points are exactly the same, but their labels are different based on θ and $f \in \{f_1, f_2\}$. When $\theta \approx 0$, f_1 appears to have better performance than f_2 in terms of the distances between points with unequal labels. The situation is reversed when $\theta \approx \frac{\pi}{2}$. This shows that the performance for a fixed decoding function can vary widely with θ even when both $|h_A|$ and $|h_B|$ are large.

As illustrated in Fig. 4, each $f \in \mathcal{F}$ induces a mapping $\mathcal{M}_{f,R}$ between address vectors $\underline{x}_{f,R}$ and constellation points $q_R \in \mathcal{Q}_R$ similar to $\mathcal{M}$ for the point to point case. The relay is only interested in decoding $\mathbf{X}_{f,R}$, which will have ℓ rows. Thus $\mathcal{M}_{f,R}$ forms a one-to-many map from length ℓ binary address vectors to constellation points. Let $\mathcal{S} \subseteq \{1, ..., \ell\}$ be the subset of elements from $\underline{x}_{f,R}$ which are fixed. Then, let $\mathcal{X}_{\{x_{f,R}^k|k\in\mathcal{S}\}} \subseteq \mathbb{F}_2^\ell$ be the subset of $\underline{x}_{f,R}$'s with the same values for all points in $\mathcal{S}$. For a given $f \in \mathcal{F}$, the output of $\mathcal{M}_{f,R}$ is

$$\mathcal{M}_{f,R}(\{x_{f,R}^k|k \in \mathcal{S}\}) = \{q_R \in \mathcal{Q}_R | q_R = h_A\mathcal{M}(\underline{x}_A) + h_B\mathcal{M}(\underline{x}_B), \ f(\underline{x}_A, \underline{x}_B) \in \mathcal{X}_{\{x_{f,R}^k|k\in\mathcal{S}\}}\}. \tag{12}$$

For the example in Fig. 4, $\mathcal{M}_{f,R}(11)$ would return the four constellation points labeled 11 in each figure. $\mathcal{M}_{f,R}(1?)$ would return the eight constellation points in the union $\mathcal{M}_{f,R}(11) \cup \mathcal{M}_{f,R}(10)$.

In order for nodes A and B to be able to unambiguously decode their desired messages, the authors in [11] show

that f must satisfy

$$f(\underline{x}_A, \underline{x}_B) \neq f(\underline{x}'_A, \underline{x}_B) \ \forall \ \underline{x}_A \neq \underline{x}'_A \text{ and } \underline{x}_B$$

$$f(\underline{x}_A, \underline{x}_B) \neq f(\underline{x}_A, \underline{x}'_B) \ \forall \ \underline{x}_B \neq \underline{x}'_B \text{ and } \underline{x}_A. \tag{13}$$

We call functions that satisfy this property unambiguous.

Lemma 1: For any $\mathbf{D}_A, \mathbf{D}_B \in \mathcal{D}$, a decoding function

$$f(\underline{x}_A, \underline{x}_B) = [\mathbf{D}_A \mathbf{D}_B] \begin{bmatrix} \underline{x}_A \\ \underline{x}_B \end{bmatrix} \tag{14}$$

is unambiguous.

Proof: The proof follows from the invertibility of $\mathbf{D}_A$ and $\mathbf{D}_B$. For some $\underline{x}_A$, suppose that there exists $\underline{x}_B \neq \underline{x}'_B$ so that

$$[\mathbf{D}_A \mathbf{D}_B] \begin{bmatrix} \underline{x}_A \\ \underline{x}_B \end{bmatrix} = [\mathbf{D}_A \mathbf{D}_B] \begin{bmatrix} \underline{x}_A \\ \underline{x}'_B \end{bmatrix}.$$

This can be written as

$$\mathbf{D}_A \underline{x}_A \oplus \mathbf{D}_B \underline{x}_B = \mathbf{D}_A \underline{x}_A \oplus \mathbf{D}_B \underline{x}'_B$$

$$\mathbf{D}_A \underline{x}_A \oplus \mathbf{D}_A \underline{x}_A \oplus \mathbf{D}_B \underline{x}_B = \mathbf{D}_A \underline{x}_A \oplus \mathbf{D}_A \underline{x}_A \oplus \mathbf{D}_B \underline{x}'_B$$

$$\mathbf{D}_B \underline{x}_B = \mathbf{D}_B \underline{x}'_B$$

$$\mathbf{D}_B^{-1} \mathbf{D}_B \underline{x}_B = \mathbf{D}_B^{-1} \mathbf{D}_B \underline{x}'_B$$

$$\underline{x}_B = \underline{x}'_B$$

which is a contradiction. ∎

IV. ACHIEVABLE INFORMATION RATES

A. Achievable Rate for a Given Function

For a given f and fixed channel gains h_A and h_B the achievable rate region is given by the following theorem. This theorem is the key contribution of this paper.

Theorem 1: Choose some fixed $\mathbf{D}_A, \mathbf{D}_B \in \mathcal{D}$ and define

$$\underline{x}_{f,R} = f(\underline{x}_A, \underline{x}_B) = [\mathbf{D}_A \mathbf{D}_B] \begin{bmatrix} \underline{x}_A \\ \underline{x}_B \end{bmatrix}. \tag{15}$$

Choose a subset $\mathcal{S} \subseteq \{1, ..., \ell\}$ and define $\overline{\mathcal{S}} = \{1, ..., \ell\} \setminus \mathcal{S}$. Divide $\mathcal{S}$ into p non-empty disjoint subsets $\mathcal{S}_1, ..., \mathcal{S}_p$ so that $\bigcup_{i=1}^p \mathcal{S}_i = \mathcal{S}$. Let Z_i, $i \in \{1, ..., p\}$ define p i.i.d. Bernoulli random variables with parameter $\frac{1}{2}$. At last, let each row of $\mathbf{X}_A$ and $\mathbf{X}_B$ be encoded using a different coset of the same linear code $\mathcal{C}$. Then there exists a linear code $\mathcal{C}$ of rate $\mathcal{R}$ for which the relay can reliably decode $\mathbf{X}_{f,R}$ as long as $\mathcal{R}$ satisfies

$$\mathcal{R} < \min_{\mathcal{S}, \overline{\mathcal{S}}, \mathcal{S}_1, ..., \mathcal{S}_p} \frac{1}{p} I(Y_R; \{X_{f,R}^k | k \in \mathcal{S}\} | \{X_{f,R}^k | k \in \overline{\mathcal{S}}\}, \{X_{f,R}^k \oplus Z_i | k \in \mathcal{S}_i\} \ \forall \ i \in \{1, ..., p\}). \tag{16}$$

For the special case when $\ell = 2$, the set of bounds described by (16) are equivalent to

$$\mathcal{R} < min\{\frac{1}{2}I(Y_R; X_R^1, X_R^2),\ I(Y_R; X_R^1|X_R^2),\ I(Y_R; X_R^2|X_R^1),\ I(Y_R; X_R^1, X_R^2|X_R^1 \oplus Z_1, X_R^2 \oplus Z_1)\}. \quad (17)$$

Note that

$$I(Y_R; X_R^1, X_R^2|X_R^1 \oplus Z_1, X_R^2 \oplus Z_1) = I(Y_R; X_R^1, X_R^2|X_R^1 \oplus X_R^2).$$

That is, $\{X_R^1 \oplus Z_1, X_R^2 \oplus Z_1\}$ and $\{X_R^1 \oplus X_R^2\}$ carry the same information about X_R^1 and X_R^2.

Proof: The detailed proof is provided in the Appendix. However, the key steps in the proof are outlined below.

Our proof uses the standard approach of deriving upper bounds on the probability of error for a joint typicality decoder averaged over a carefully chosen ensemble of codes. The ensemble considered here is the ensemble obtained by using random cosets of the *same linear* code for the different signaling levels in the multilevel coding scheme. The linear code is chosen from the ensemble of linear codes with randomly chosen entries in the generator matrix. *The use of the same linear code in each level is an important ingredient in our proposed scheme since we allow the relay to freely take linear combinations of codewords from different signaling levels. However, this is also what complicates the proof.* The ensemble used here is different from the often used ensemble of random coset codes used at each level in the multilevel coding scheme since the latter ensemble allows for independently chosen codes at each level. While the latter ensemble has been used widely to obtain achievable rates for MLC for the point to point channel and the multiple access channel, the former ensemble has not been analyzed in detail in the literature. The key contribution of our proof in the Appendix is to derive the achievable rates with the former ensemble with identical linear codes at each level.

This can be accomplished since the use of the same linear code at each level ensures that for each $f \in \mathcal{F}$, $\mathcal{C}_{f,R}^k$, $k \in \{1, ..., \ell\}$ is a member of the ensemble used at the transmitters. The main complication that arises from this is that the pairwise independence assertion that is required in typical channel coding proofs [15] does not hold for certain classes of error events. Particularly, it is possible for the relay to correctly decode some rows of $\mathbf{X}_{f,R}$ while others may be in error. We handle this by splitting the union bound for error probability into separate classes of error events which are conditionally pairwise independent.

The bound for the $\ell = 2$ case can be derived by letting $\mathcal{S}, \overline{\mathcal{S}}, \mathcal{S}_1, \mathcal{S}_2 \subseteq \{1, 2\}$ take the following values respectively.

$$\{\mathcal{S} = \{1, 2\}, \overline{\mathcal{S}} = \emptyset, \mathcal{S}_1 = \{1\}, \mathcal{S}_2 = \{2\}\}$$

$$\{\mathcal{S} = \{1\}, \overline{\mathcal{S}} = \{2\}, \mathcal{S}_1 = \{1\}\}$$

$$\{\mathcal{S} = \{2\}, \overline{\mathcal{S}} = \{1\}, \mathcal{S}_1 = \{2\}\}$$

$$\{\mathcal{S} = \{1, 2\}, \overline{\mathcal{S}} = \emptyset, \mathcal{S}_1 = \{1, 2\}\}. \quad (18)$$

Notice that the first three terms in (17) are also required by the proof for multilevel coding for the point to point channel. The last bound is a result of the requirement that each signaling level uses a coset of the same linear code. It would be required for the point to point case as well if the same codes were used at each level. ∎

It should be noted that the steps of the proof for theorem 1 can applied almost unaltered to the problem of finding the achievable rate for decode-and-forward if nodes A and B transmit using different cosets of the same linear codes at each level. In a decode-and-forward scheme, the relay attempts to reliably decode the messages transmitted from node A and B and then broadcasts a function of the received messages to nodes A and B. With a slight change to the channel model, the proof of theorem 1 can be applied to the problem of recovering the 2ℓ coset codewords which form the rows of

$$\mathbf{X}_{AB} = \begin{bmatrix} \mathbf{X}_A \\ \mathbf{X}_B \end{bmatrix}. \tag{19}$$

Dividing the set $\{1, ..., 2\ell\}$ into subsets $\mathcal{S}, \overline{\mathcal{S}}, \mathcal{S}_1, ..., \mathcal{S}_p$ as in theorem 1, we can show that $\mathbf{X}_{AB}$ can be reliably decoded as long as $\mathcal{R}$ satisfies

$$\mathcal{R} < \min_{\mathcal{S}, \overline{\mathcal{S}}, \mathcal{S}_1, ..., \mathcal{S}_p} \frac{1}{p} I(Y_R; \{X_{AB}^k | k \in \mathcal{S}\} | \{X_{AB}^k | k \in \overline{\mathcal{S}}\}, \{X_{AB}^k \oplus Z_i | k \in \mathcal{S}_i\} \ \forall \ i \in \{1, ..., p\}). \tag{20}$$

Therefore, by allowing the relay to choose between compute-and-forward and decode-and-forward, the maximum of the bounds given by (16) and (20) is achievable.

B. Universally Achievable Rate

We say that a rate $\mathcal{R}$ is universally achievable over the set $\mathcal{H} \subset \mathbb{C}^2$ if there exists a fixed linear code $\mathcal{C}$ of rate $\mathcal{R}$ and coset matrices $\mathbf{\Lambda}_A$ and $\mathbf{\Lambda}_B$ such that for every $(h_A, h_B) \in \mathcal{H}$, the relay can reliably decode $\mathbf{X}_{f,R}$ for some $f \in \mathcal{F}$. That is some $\mathbf{X}_{f,R}$ can be decoded with arbitrarily small probability of error in the usual information-theoretic sense. The main result in this section is the following theorem.

Theorem 2: For a fixed $f \in \mathcal{F}$ and (h_A, h_B), define $\mathcal{R}_f(h_A, h_B)$ as the supremum of rates satisfying (16) where $\underline{x}_{f,R} = f(\underline{x}_A, \underline{x}_B)$. For any finite set of channel gains, $\mathcal{H} \subset \mathbb{C}^2$, any rate $\mathcal{R}$ such that

$$\mathcal{R} < \min_{(h_A, h_B) \in \mathcal{H}} \max_{f \in \mathcal{F}} \mathcal{R}_f(h_A, h_B) \tag{21}$$

is universally achievable.

Proof: For a fixed finite $\mathcal{H} \subset \mathbb{C}^2$ and set of decoding functions $\mathcal{F}$, define $\mathcal{R}'$ as the supremum of rates satisfying (21). Define $\delta > 0$ as the acceptable probability of error for a finite length code and choose a fixed $\mathcal{R} < \mathcal{R}'$.

We will first consider an arbitrary $(h_A, h_B) \in \mathcal{H}$ and $f \in \mathcal{F}$ such that $\mathcal{R} < \mathcal{R}_f(h_A, h_B)$. Define Ω^N as the set of coset codes of the form $\{\mathcal{C}, \mathbf{\Lambda}_A, \mathbf{\Lambda}_B\}$ which have length N. Thus, by increasing the value of N we form a sequence of ensembles of coset codes Ω. Define $P(Err|\Omega^N)$ as the ensemble average probability of decoding error for the ensemble Ω^N. Define $P(Err|\mathcal{C}, \mathbf{\Lambda}_A, \mathbf{\Lambda}_B)$ as the probability of decoding error for a specific coset code.

Define $\Omega_{bad}^N \subset \Omega^N$ as

$$\Omega_{bad}^N = \{\mathcal{C}, \mathbf{\Lambda}_A, \mathbf{\Lambda}_B \in \Omega^N | P(Err|\mathcal{C}, \mathbf{\Lambda}_A, \mathbf{\Lambda}_B) \geq \delta\}. \tag{22}$$

Then let $\Omega_{good}^N = \Omega^N \setminus \Omega_{bad}^N$. Define $P(bad|N) = \frac{|\Omega_{bad}^N|}{|\Omega^N|}$ and $P(good|N) = \frac{|\Omega_{good}^N|}{|\Omega^N|}$ as the probability that a bad

or good code is selected uniformly at random from Ω^N respectively. We know that

$$P(Err|\Omega^N) = P(bad|N)P(Err|\Omega^N_{bad}) + P(good|N)P(Err|\Omega^N_{good})$$

$$\geq P(bad|N)\delta + P(good|N)P(Err|\Omega^N_{good})$$

$$\geq P(bad|N)\delta.$$

The proof of theorem 1 relies on showing that $\lim_{N \to \infty} P(Err|\Omega^N) = 0$. Therefore, there exists some N_0 such that for any $N > N_0$,

$$P(bad|N)\delta \leq P(Err|\Omega^N) < \frac{\delta}{\tau} \;\Rightarrow\; P(bad|N) < \frac{1}{\tau}$$

for some finite $\tau > 2|\mathcal{H}|$. This means that $|\Omega^N_{bad}| < \frac{|\Omega^N|}{\tau}$. Note that choosing $\tau > 2|\mathcal{H}|$ is arbitrary but ensures that τ will be "large enough" to complete the proof.

We want to show the existence of some fixed $\{\mathcal{C}, \boldsymbol{\Lambda}_A, \boldsymbol{\Lambda}_B\} \in \Omega^N$ such that for every $(h_A, h_B) \in \mathcal{H}$ there is some $f \in \mathcal{F}$ so that $P(Err|\mathcal{C}, \boldsymbol{\Lambda}_A, \boldsymbol{\Lambda}_B) < \delta$. We can apply the steps above to find a set $\Omega^N_{bad}(h_A, h_B)$ for every $(h_A, h_B) \in \mathcal{H}$. Since $|\mathcal{H}|$ is finite the largest N required by any $(h_A, h_B) \in \mathcal{H}$ must exist and be a finite integer N_{max}.

Since τ is chosen to be larger than $2|\mathcal{H}|$, the set

$$\Omega^{N_{max}} \setminus \{ \bigcup_{(h_A, h_B) \in \mathcal{H}} \Omega^{N_{max}}_{bad}(h_A, h_B)\}$$

must be non-empty because

$$\sum_{(h_A, h_B) \in \mathcal{H}} |\Omega^{N_{max}}_{bad}(h_A, h_B)| \leq |\Omega^{N_{max}}|/2.$$

Thus, since at least half of the codes are always good, there exists at least one coset code which allows reliable decoding for every $(h_A, h_B) \in \mathcal{H}$ as long as $\mathcal{R} < \mathcal{R}'$.

∎

Note that in order for this problem to be practically interesting, the set $\mathcal{H}$ should be meaningfully defined. It may seem more natural to evaluate our scheme based on the outage probability for a fixed transmission rate. We consider the universally achievable rate formulation for two reasons. First, the outage probability can be determined for the block fading channel using results from Theorems 1 and 2. Second, the universally achievable rate is useful for illustrating the flexibility of the proposed scheme to phase mismatch between nodes A and B. This is especially interesting if we consider a system where the relay is used to provide power control information to nodes A and B as in [8].

V. NUMERICAL RESULTS

A. Numerical Results for QPSK

As an example, consider the case where nodes A and B transmit symbols from a QPSK constellation with Gray Labeling. Fig. 5 shows a plot of the achievable information rate $\ell\mathcal{R}_f(h_A, h_B)$ as given in (17) for each function
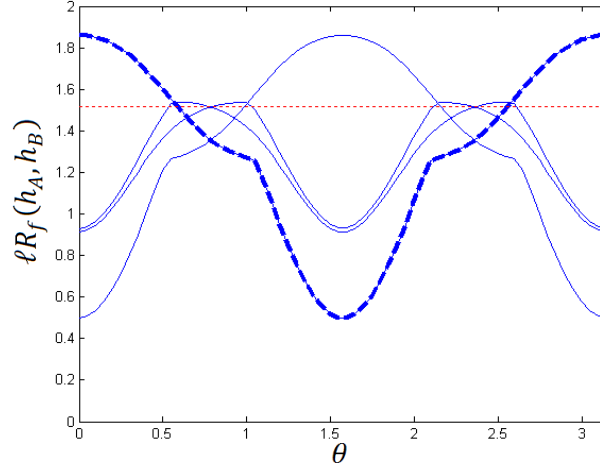
Fig. 5. $\ell\mathcal{R}_f(h_A, h_B)$ vs. θ for each $f \in \mathcal{F}$.
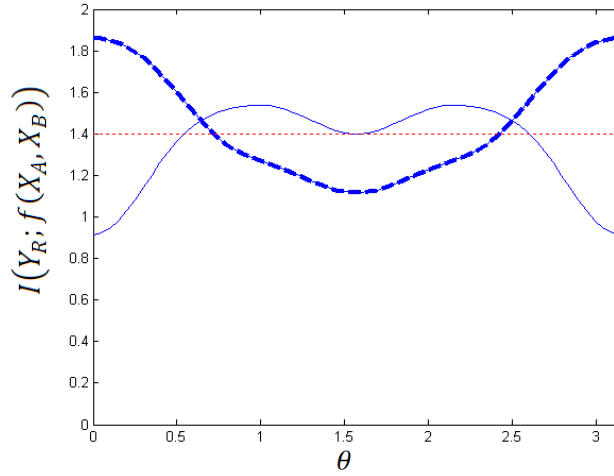

Fig. 6. $I(Y_R; f(X_A, X_B))$ vs. θ for each $f \in \mathcal{F}_{GF4}$.

$f \in \mathcal{F}$ dependent on the phase difference $\theta = \theta_A - \theta_B$ for an SNR of 7 dB. $\mathcal{H}$ is the set of channel gains

$$\mathcal{H} = \{(h_A, h_B)|h_A = e^{j\theta_A}, h_B = e^{j\theta_B}\} \tag{23}$$

where $\theta_A, \theta_B \in \{0, \frac{\pi}{m}, ..., 2\pi\}$ for a finite integer m. Thus $|\mathcal{H}|$ is finite but approximates the selection of any value of θ_A and θ_B arbitrarily closely.

The dotted line indicates the universally achievable rate in bits per complex symbol for the proposed scheme which satisfies Theorem 2 for $\mathcal{H}$. Note that different functions provide the best performance for different values of θ which reiterates the substantial benefit of decoding adaptively. Notice that a small increase in rate makes reliable decoding impossible for any $f \in \mathcal{F}$ for a significant range of θ; however, there are many $(h_A, h_B) \notin \mathcal{H}$ such that

$\exists f \in \mathcal{F}$ for which reliable decoding is possible.

B. Coding over $GF(4)$

It is interesting to use this QPSK example to compare our MLC scheme to the case where nodes A and B encode their a data using a linear code $\mathcal{C}_{GF4}$ over $\mathbb{F}_4$ of rate $\mathcal{R}_{GF4}$. The relay uses the set of decoding functions $\mathcal{F}_{GF4}$ corresponding to linear combinations of codewords in $\mathbb{F}_4$ of the form

$$\underline{v}_R = f(\underline{v}_A, \underline{v}_B) = \alpha \underline{v}_A \oplus \beta \underline{v}_B, \ \alpha, \beta \in \mathbb{F}_4 \backslash \{0\}. \tag{24}$$

Node A can decode $\underline{v}_B$ from $\underline{v}_A$ and $\underline{v}_R$ by

$$\underline{v}_B = \beta^{-1}(\alpha \underline{v}_A) \oplus \underline{v}_R. \tag{25}$$

Node B can recover $\underline{v}_A$ similarly. The relay should be able to decode $\underline{v}_R$ reliably as long as there exists some $f \in \mathcal{F}_{GF4}$ for which

$$\mathcal{R}_{GF4} < I(Y_R; f(X_A, X_B)). \tag{26}$$

The value of $I(Y_R; f(X_A, X_B))$ for each possible $f \in \mathcal{F}_{GF4}$ is plotted as a function of θ in Fig. 6 with an SNR of 7 dB. Again the dotted line represents universally achievable rate for the $\mathcal{H}$ in (23).

C. Comparison of Proposed Techniques

These numerical results illustrate that the proposed MLC scheme facilitates better decoding flexibility at the relay than coding over $\mathbb{F}_4$ for this example. In fact, in an analysis of these functions based on the labeling of points in $\mathcal{Q}_R$, it can be seen that $\mathcal{F}_{GF4} \subset \mathcal{F}$. However, this improved flexibility comes at the cost of additional rate constraints on each $f \in \mathcal{F}$. The thick dashed line in Figs. 5 and 6 represents the rate which is achievable if the relay decodes using some f which is equivalent to the componentwise xor operation for multilevel coding or finite field addition for $\mathbb{F}_4$. The difference between these curves illustrates the effects of the additional rate constraints imposed by (16). In Fig. 5 the last term $I(Y_R; X_R^1, X_R^2 | X_R^1 \oplus Z_1, X_R^2 \oplus Z_1)$ in (17) is dominant if $\theta \approx \frac{\pi}{2}$ for determining the achievable rate for this function. In Fig. 6 we see that this term does not need to be satisfied if nodes A and B use a linear code in $\mathbb{F}_4$.

The universally achievable rate for the $\mathcal{H}$ in (23) (i.e. the constant value given by the dotted line in Figs. 5 and 6) is plotted as a function of SNR in Fig. 7 for the cases where the relay uses $\mathcal{F}$ or $\mathcal{F}_{GF4}$. This value asymptotically approaches 1.5 bits per symbol for coding over $\mathbb{F}_4$. From Fig. 6, this appears to occur because $\mathcal{F}_{GF4}$ does not provide the relay with a decoding function which works well when $\theta \approx \frac{\pi}{2}$. This represents an extreme case, because the event $|h_A| = |h_B|$ occurs with probability zero for many random fading processes. However, this illustrates that for PLNC it is possible for the universally achievable rate to be limited by specific $(h_A, h_B) \in \mathcal{H}$ even if each $|h_A|$ and $|h_B|$ is large.
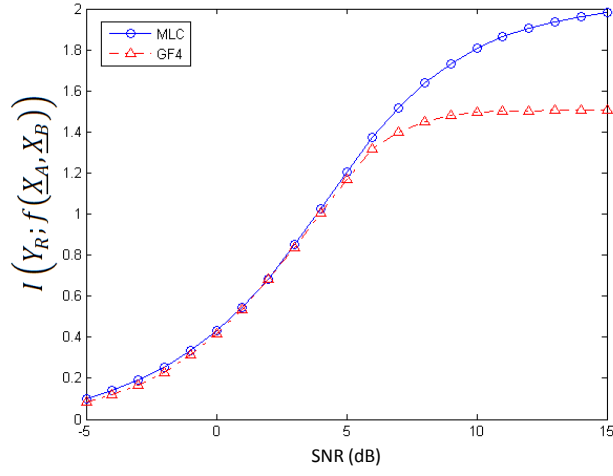
Fig. 7. Universally achievable rates vs. SNR(dB) for proposed relaying techniques
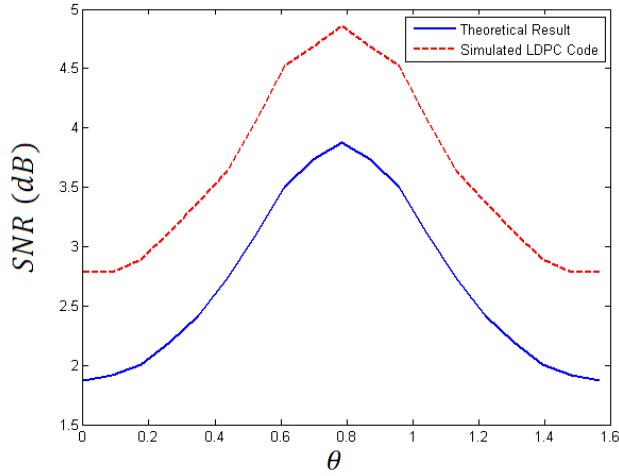

Fig. 8. Required SNR(dB) vs. θ to reliably decode a rate $\frac{1}{2}$ code

D. Simulation Results

To corroborate these theoretical results, we simulated the performance of a regular (3,6) low density parity check (LDPC) code. In Fig. 8 the required SNR for a rate $\frac{1}{2}$ code is plotted as a function of $\theta = \theta_A - \theta_B$ for the case where $(h_A, h_B) \in \mathcal{H}$ for the $\mathcal{H}$ in (23). The solid curve represents the theoretically required SNR as determined by (16). The dashed red curve represents the SNR for which zero bit errors occurred during 200 simulations of a length 10^5 code for each tested θ. For a point to point Gaussian channel using binary phase shift keying, it has been shown in [16] that the required SNR for a (3,6) LDPC code with iterative decoding is about 1dB away from the Shannon limit for the same channel and modulation format. In Fig. 8, we see that this trend appears to hold

for our scheme as well. We leave more rigorous testing for future work. Note that to achieve the theoretical limit imposed by (16) using structured codes, it will be necessary to design coding schemes which universally achieve the capacity for many channel conditions. It appears that the class of spatially coupled LDPC codes would be a good choice for this [17].

VI. CONCLUDING REMARKS

In this paper, we have proposed a coding scheme based on MLC for compute and forward or PLNC for the case when the channel is perfectly estimated at each receiver but unknown to each transmitter. We showed that MLC allows for decoding of a set of functions of the transmitted messages and the relay can choose one function from this set depending on the channel coefficients. In Theorem 1, we obtained an achievable rate for a fixed decoding function and channel realizations. In Theorem 2, we obtained a numerically computable expression for the universally achievable information rate over a set of channel realizations. Numerical results for QPSK suggest that the proposed scheme significantly outperforms the use of a fixed decoding function with binary linear codes and is better than using linear codes over $\mathbb{F}_4$.

APPENDIX

PROOF OF THEOREM 1

Theorem 1 states that for a fixed $f \in \mathcal{F}$, if each $\mathcal{C}_A^1, ..., \mathcal{C}_A^\ell$ and $\mathcal{C}_B^1, ..., \mathcal{C}_B^\ell$ is a different coset of the same linear code $\mathcal{C}$ of rate $\mathcal{R}$, then there exists some $\mathcal{C}$ for which the relay can reliably decode $\mathbf{X}_{f,R}$ for a suitably chosen $\mathcal{R}$.

A. Additional Notation

A few definitions only necessary for this proof have been omitted from the main text but are included here for clarity.

We refer to the noiseless observed sequence at the relay as

$$\underline{q}_R = h_A \underline{s}_A + h_B \underline{s}_B. \tag{27}$$

Where $\underline{q}_R \in \mathcal{Q}_R^N$. Thus the relay observes the noisy observations

$$\underline{y}_R = h_A \underline{s}_A + h_B \underline{s}_B + \underline{w}_R = \underline{q}_R + \underline{w}_R. \tag{28}$$

When it is necessary to refer to variables associated with different messages, we will refer to variables like $\mathbf{U}_{j,A}$ by the integer $j \in \{0, ..., 2^{K\ell} - 1\}$, whose binary expansion is given by node A's unparsed message $\underline{u}_A$. We assume that nodes A and B encode $\mathbf{U}_{j_A,A}$ and $\mathbf{U}_{j_B,B}$ for transmission, and that the relay observes the noisy samples corresponding to $\mathbf{X}_{j_R,f,R}$. Note that the index j_R of the desired message is a function of j_A, j_B, and f.

The relay will attempt to reliably decode $\mathbf{X}_{j_R,f,R}$ from $\underline{y}_R$ using a joint typicality decoder. Thus the decoder declares an error if either $\mathbf{X}_{j_R,f,R}$ is not jointly typical with $\underline{y}_R$ or if some incorrect message $\mathbf{X}_{j,f,R}$, $j \neq j_R$ is jointly typical with $\underline{y}_R$. We derive an upper bound on the error probability for this decoder over the ensemble of random coset codes. Specifically, let the elements of $\mathbf{G}$, $\mathbf{\Lambda}_A$, and $\mathbf{\Lambda}_B$ be i.i.d. Bernoulli random variables with parameter $\frac{1}{2}$.

B. Pairwise Independence of Codewords

Here we provide a brief analysis of the ensemble of coset codes used by nodes A and B and observed by the relay. The following lemmas are stated as they pertain to a nameless encoder to simplify notation. Both lemmas appear as part of the proof of Gallager's Coding Theorem for Random Parity Check Codes [15]. We include these proofs because the intuition behind some of the steps is used for other parts of the proof of Theorem 1.

Lemma 2: Let each element of $\mathbf{G}$ and $\underline{\lambda}^k$ be i.i.d. Bernoulli random variables with parameter $\frac{1}{2}$. Then we have

$$P(\underline{V}_j^k = \underline{v}_j^k) = \frac{1}{2^N} \ \forall \ \underline{v}_j^k \in \mathbb{F}_2^N. \tag{29}$$

That is, the codeword $\underline{v}_j^k \in \mathbb{F}_2^N$ associated with message vector $\underline{u}_j^k \in \mathbb{F}_2^K$ can take any value with uniform probability over the ensemble of random coset codes.

Proof: For a fixed $\mathbf{G}$ and $\underline{u}_j^k$, the output of the linear encoder $\underline{\gamma}_j^k = \underline{u}_j^k \mathbf{G}$ must take some value in $\mathbb{F}_2^N$. Since $\underline{\lambda}^k$ can take any value with equal probability we have

$$P(\underline{V}_j^k = \underline{v}_j^k) = P(\underline{\Lambda}^k = \underline{v}_j^k \oplus \underline{\gamma}_j^k) = \frac{1}{2^N} \ \forall \ \underline{v}_j^k \in \mathbb{F}_2^N. \tag{30}$$

∎

Lemma 3: Let each element of $\mathbf{G}$ and $\underline{\lambda}^k$ be i.i.d. Bernoulli random variables with parameter $\frac{1}{2}$. Then for any $j' \neq j$ for which $\underline{u}_j^k \neq \underline{u}_{j'}^k$, we have

$$P(\underline{V}_j^k = \underline{v}_j^k, \underline{V}_{j'}^k = \underline{v}_{j'}^k) = P(\underline{V}_j^k = \underline{v}_j^k) P(\underline{V}_{j'}^k = \underline{v}_{j'}^k)$$
$$= \frac{1}{2^{2N}} \ \forall \ \underline{v}_j^k, \underline{v}_{j'}^k \in \mathbb{F}_2^N. \tag{31}$$

That is, the codewords $\underline{v}_j^k$ and $\underline{v}_{j'}^k$ associated with $\underline{u}_j^k$ and $\underline{u}_{j'}^k$ respectively are pairwise independent and uniformly distributed over $\mathbb{F}_2^N$.

Proof: Suppose that $\underline{u}_j^k$ and $\underline{u}_{j'}^k$ differ in the m_{th} position, and let $\underline{g}_i$, $i \in \{1, ..., K\}$ refer to the i_{th} row of $\mathbf{G}$. Then for any set of rows

$$\underline{g}_1, ..., \underline{g}_{m-1}, \underline{g}_{m+1}, ..., \underline{g}_K$$

there is some $\underline{g}_m$ which gives $\underline{v}_j^k \oplus \underline{v}_{j'}^k = \underline{\gamma}_j^k \oplus \underline{\gamma}_{j'}^k$, any fixed value. By the construction of $\mathbf{G}$ and Lemma 2, $\underline{g}_m$ and $\underline{v}_j^k$ can take any value with uniform probability. We can conclude that

$$P(\underline{V}_j^k = \underline{v}_j^k, \underline{V}_{j'}^k = \underline{v}_{j'}^k | \underline{u}_j^k \neq \underline{u}_{j'}^k)$$
$$= P(\underline{V}_j^k = \underline{v}_j^k | \underline{u}_j^k \neq \underline{u}_{j'}^k) P(\underline{V}_{j'}^k = \underline{v}_{j'}^k | \underline{v}_j^k, \underline{u}_j^k \neq \underline{u}_{j'}^k)$$
$$= P(\underline{\Lambda}^k = \underline{v}_j^k \oplus \underline{\gamma}_j^k | \underline{u}_j^k \neq \underline{u}_{j'}^k) P(\underline{G}_m = \underline{v}_{j'}^k \oplus \underline{v}_j^k | \underline{v}_j^k, \underline{u}_j^k \neq \underline{u}_{j'}^k)$$
$$= \frac{1}{2^N} \frac{1}{2^N} = \frac{1}{2^{2N}} \ \forall \ \underline{v}_j^k, \underline{v}_{j'}^k \in \mathbb{F}_2^N. \tag{32}$$

∎

The key idea behind each proof is the same. In Lemma 2, we see that the uniform distribution of $\underline{\lambda}^k$ implies the uniform distribution of $\underline{v}_j^k$. In Lemma 3, we see that the uniform distribution of $\mathbf{G}$ implies the pairwise independence of codewords corresponding to distinct messages.

C. Distribution of Received Signal

The last step before deriving upper bounds on the probability of decoding error is to derive the distribution of the signal received by the relay over the ensemble of codes.

Lemma 4: Let $\mathbf{T} \in \mathbb{F}_2^{m \times 2\ell}$ be a matrix of rank m, and let $\underline{x}_{\mathbf{T}}[n] \in \mathbb{F}_2^m$ be defined by

$$\underline{x}_{\mathbf{T}}[n] = \mathbf{T} \begin{bmatrix} \underline{x}_{j_A,A}[n] \\ \underline{x}_{j_B,B}[n] \end{bmatrix}. \tag{33}$$

For a fixed $\mathbf{T}$, define the set $\mathcal{Q}_{\underline{x}_{\mathbf{T}},R}$ by

$$\mathcal{Q}_{\underline{x}_{\mathbf{T}},R} = \{q_R \in \mathcal{Q}_R | q_R = h_A \mathcal{M}(\underline{x}_A) + h_B \mathcal{M}(\underline{x}_B) \text{ for some } \underline{x}_A, \underline{x}_B \in \mathbb{F}_2^\ell \text{ such that } \underline{x}_{\mathbf{T}}[n] = \mathbf{T} \begin{bmatrix} x_A \\ x_B \end{bmatrix}\}. \tag{34}$$

Then the distribution of $\underline{q}_R[n]$ conditioned on $\underline{x}_{\mathbf{T}}[n]$ is given by

$$P(\underline{Q}_R[n] = q_R | \underline{x}_{\mathbf{T}}[n]) = \frac{1}{2^{2\ell-m}} \ \forall \ q_R \in \mathcal{Q}_{\underline{x}_{\mathbf{T}}[n],R}. \tag{35}$$

Proof: For a given $\mathbf{U}_{j_A,A}$ and $\mathbf{U}_{j_B,B}$, each $\underline{v}_{j_A,A}^k, \underline{v}_{j_B,B}^k$, $k \in \{1,...,\ell\}$ is uniformly distributed in $\mathbb{F}_2^N$ by Lemma 2. Thus over the ensemble of codes, each element of $\mathbf{X}_{j_A,A}$ and $\mathbf{X}_{j_B,B}$ is an i.i.d. Bernoulli random variable with parameter $\frac{1}{2}$. By (27), we have

$$\underline{q}_R[n] = h_A \mathcal{M}(\underline{x}_{j_A,A}[n]) + h_B \mathcal{M}(\underline{x}_{j_B,B}[n]). \tag{36}$$

Thus for fixed channel gains, $\underline{q}_R[n]$ is a bijective function of $\underline{x}_{j_A,A}[n]$ and $\underline{x}_{j_B,B}[n]$. For a given $\underline{x}_{\mathbf{T}}[n]$ we have,

$$P(\underline{X}_{j_A,A}[n] = \underline{x}_A, \underline{X}_{j_B,B}[n] = \underline{x}_B | \underline{x}_{\mathbf{T}}[n]) = \frac{1}{2^{2\ell-m}} \ \forall \ \underline{x}_A, \underline{x}_B \in \mathbb{F}_2^\ell \text{ such that } \underline{x}_{\mathbf{T}}[n] = \mathbf{T} \begin{bmatrix} x_A \\ x_B \end{bmatrix}.$$

That is the distribution of 2ℓ i.i.d. Bernoulli random variables with parameter $\frac{1}{2}$ conditioned on m linear combinations of these variables is uniform over the space of outcomes satisfying the linear constraints. The result follows because $\underline{q}_R[n]$ is a function of the address vectors. $\blacksquare$

By (28), the conditional distribution of $\underline{y}_R[n]$ on $\{x_{j_R,f,R}^k[n] | k \in \mathcal{S}\}$ is given by

$$P(y_R[n]|\{x_{j_R,f,R}^k[n]|k \in \mathcal{S}\}) = \frac{1}{2^{2\ell-|\mathcal{S}|}} \sum_{q_R \in \mathcal{M}_{f,R}(\{x_{j_R,f,R}^k[n]|k \in \mathcal{S}\})} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{|y_R[n]-q_R|^2}{2\sigma^2}} \tag{37}$$

where $\mathcal{M}_{f,R}$ is the mapping function at the relay induced by f and the channel conditions.

D. Analysis of Error Probability

The relay uses a joint typicality decoder to decode $\mathbf{X}_{j_R,f,R}$ from $\underline{y}_R$. For some fixed $\epsilon > 0$, define $\mathcal{A}_\epsilon^N$ as the set of $(\mathbf{X}_{j,f,R}, \underline{y}_R)$ pairs which satisfy the definition of joint typicality given in [18]. The set $\mathcal{A}_\epsilon^N$ is referred to as the jointly typical set. Let the event E_j, $j \in \{0,...,2^{K\ell}-1\}$ be the event $(\mathbf{X}_{j,f,R}, \underline{y}_R) \in \mathcal{A}_\epsilon^N$. The probability of error given that the codeword corresponding to j_R is observed by the relay can be expressed

$$P(Err|j_R) = P(\overline{E}_{j_R} \cup \bigcup_{\substack{j \neq j_R}}^{2^{K\ell}-1} E_j | j_R). \tag{38}$$

Applying the union bound, we get

$$P(Err|j_R) \leq P(\overline{E}_{j_R}|j_R) + \sum_{j \neq j_R}^{2^{K\ell}-1} P(E_j|j_R). \tag{39}$$

Recall that $\underline{y}_R$ is the result of the relay observing the symbol sequence associated with message j_R. Thus by the joint asymptotic equipartition property (AEP) we have that for any $\epsilon > 0$,

$$P(\overline{E}_{j_R}|j_R) < \epsilon \tag{40}$$

for sufficiently large N.

The proof of the channel coding theorem for the general discrete memoryless channel in [18] relies on upper bounding $P(E_j|j_R)$ using the joint AEP. This is not straightforward here because $\mathbf{X}_{j,f,R}$, $j \neq j_R$ and $\underline{y}_R$ are not independent with the same marginals for certain classes of error events. For example, if $\ell = 2$, we could have

$$\mathbf{U}_{j,f,R} = \mathbf{U}_{j_R,f,R} \oplus \begin{bmatrix} \underline{e}_u \\ \underline{0} \end{bmatrix}$$

$$\Rightarrow \mathbf{X}_{j,f,R} = \mathbf{X}_{j_R,f,R} \oplus \begin{bmatrix} \underline{e}_v \\ \underline{0} \end{bmatrix} \tag{41}$$

for some $\underline{e}_u \in \mathbb{F}_2^K \setminus \{0\}$ and $\underline{e}_v \in \mathbb{F}_2^N \setminus \{0\}$. This means that $\underline{v}_{j,f,R}^1 \neq \underline{v}_{j_R,f,R}^1$ but $\underline{v}_{j,f,R}^2 = \underline{v}_{j_R,f,R}^2$. Thus for this class of error events $\mathbf{X}_{j,f,R}$ and $\mathbf{X}_{j_R,f,R}$ are not pairwise independent. Note that this class of error events is handled by the proof of the coding theorem for the multiple access channel [14], [18], and [19]. In the coding theorem proof for the multiple access channel, it is possible for the receiver to correctly decode a codeword from one transmitter while making an error in decoding the codeword from a second transmitter. This has the same effect as correctly decoding the codeword on one level of a multilevel encoder while making an error in decoding the codeword transmitted on the second level.

Unfortunately, choosing to use a coset of the *same* linear codes at each bit level introduces a new class of error events of the form

$$\mathbf{U}_{j,f,R} = \mathbf{U}_{j_R,f,R} \oplus \begin{bmatrix} \underline{e}_u \\ \underline{e}_u \end{bmatrix}$$

$$\Rightarrow \mathbf{X}_{j,f,R} = \mathbf{X}_{j_R,f,R} \oplus \begin{bmatrix} \underline{e}_v \\ \underline{e}_v \end{bmatrix}. \tag{42}$$

For this class of error events, the columns of the error matrix $\mathbf{X}_{j,f,R} \oplus \mathbf{X}_{j_R,f,R}$ must be in $\{[0\ 0]^T, [1\ 1]^T\}$. *This is the key difference between our proof and the proofs for the general multiple access channel or point to point channel with multilevel coding.*

We can move forward by splitting the sum in (39) into different events for which $\mathbf{X}_{j,f,R}$ and $\mathbf{X}_{j_R,f,R}$ are conditionally pairwise independent. Define a set of $p \leq \ell$ disjoint subsets $\mathcal{S}_1, ..., \mathcal{S}_p \subseteq \{1, ..., \ell\}$. Let t_m be the smallest element of $\mathcal{S}_m$, and define the sets $\mathcal{T} = \{t_1, ..., t_p\}$, $\mathcal{S} = \mathcal{S}_1 \cup ... \cup \mathcal{S}_p$, and $\overline{\mathcal{S}} = \{1, ..., \ell\} \setminus \mathcal{S}$. For each

set of subsets, define an index set $\mathcal{J}_{\mathcal{S}_1,\dots,\mathcal{S}_p}$ given by

$$\mathcal{J}_{\mathcal{S}_1,\dots,\mathcal{S}_p} = \left\{ j \,\middle|\, \underline{u}_{j,f,R}^k = \begin{cases} \underline{u}_{j_R,f,R}^k \oplus \underline{e}_{u,1} & , k \in \mathcal{S}_1 \\ \vdots & \\ \underline{u}_{j_R,f,R}^k \oplus \underline{e}_{u,p} & , k \in \mathcal{S}_p \\ \underline{u}_{j_R,f,R}^k & , k \notin \mathcal{S} \end{cases} \right\}. \tag{43}$$

Here each message error vector, $\underline{e}_{u,i} \in \mathbb{F}_2^K \backslash \{\underline{0}\} \ \forall \ i \in \{1,\dots,p\}$ satisfies $\underline{e}_{u,i} \neq \underline{e}_{u,i'} \ \forall \ i \neq i'$. For the sake of simplicity, we will complete the analysis of error probability for the case where $\ell = 2$, and then extend the results to a general ℓ.

Case $\{\ell = 2\}$: If $\ell = 2$, the subsets in (43) can be written as

$$\mathcal{J}_{\{1\}} = \left\{ j \,\middle|\, \mathbf{U}_{j,f,R} = \mathbf{U}_{j_R,f,R} \oplus \begin{bmatrix} \underline{e}_{u,1} \\ \underline{0} \end{bmatrix} \right\}$$

$$\mathcal{J}_{\{2\}} = \left\{ j \,\middle|\, \mathbf{U}_{j,f,R} = \mathbf{U}_{j_R,f,R} \oplus \begin{bmatrix} \underline{0} \\ \underline{e}_{u,1} \end{bmatrix} \right\}$$

$$\mathcal{J}_{\{1,2\}} = \left\{ j \,\middle|\, \mathbf{U}_{j,f,R} = \mathbf{U}_{j_R,f,R} \oplus \begin{bmatrix} \underline{e}_{u,1} \\ \underline{e}_{u,1} \end{bmatrix} \right\}$$

$$\mathcal{J}_{\{1\}\{2\}} = \left\{ j \,\middle|\, \mathbf{U}_{j,f,R} = \mathbf{U}_{j_R,f,R} \oplus \begin{bmatrix} \underline{e}_{u,1} \\ \underline{e}_{u,2} \end{bmatrix} \right\}.$$

These subsets are disjoint and cover each error event, so

$$\mathcal{J}_{\{1\}} \cup \mathcal{J}_{\{2\}} \cup \mathcal{J}_{\{1,2\}} \cup \mathcal{J}_{\{1\}\{2\}} = \{0,\dots,2^{2K}-1\} \backslash \{j_R\}.$$

Therefore, the union bound on the probability of error for $\ell = 2$ can be written as

$$P(Err|j_R) \leq P(\overline{E}_{j_R}|j_R) + \sum_{j \in \mathcal{J}_{\{1\}}} P(E_j|j_R, j \in \mathcal{J}_{\{1\}})$$

$$+ \sum_{j \in \mathcal{J}_{\{2\}}} P(E_j|j_R, j \in \mathcal{J}_{\{2\}})$$

$$+ \sum_{j \in \mathcal{J}_{\{1,2\}}} P(E_j|j_R, j \in \mathcal{J}_{\{1,2\}})$$

$$+ \sum_{j \in \mathcal{J}_{\{1\}\{2\}}} P(E_j|j_R, j \in \mathcal{J}_{\{1\}\{2\}}). \tag{44}$$

We define $\underline{e}_{v,i} = \underline{e}_{u,i}\mathbf{G}, \ i \in \{1,2\}$ as the codeword error vector associated with subset $\mathcal{S}_i$. The subscript u or v is used to differentiate between the message error vector and codeword error vector respectively. Over the ensemble of codes, each $\underline{e}_{v,i}$ is uniformly distributed in $\mathbb{F}_2^N$, and codeword error vectors $\underline{e}_{v,i}, \underline{e}_{v,j}, \ i \neq j$ are pairwise independent and identically distributed. These facts can be shown using steps similar to the proofs of Lemmas 2 and 3.

If $\{j \in \mathcal{J}_{\{1\}}\}$ is given, then we know that

$$\mathbf{X}_{j,f,R} = \mathbf{X}_{j_R,f,R} \oplus \begin{bmatrix} \underline{e}_{v,1} \\ \underline{0} \end{bmatrix}.$$

By Lemmas 2 and 3, $\underline{v}^1_{j_R,f,R}$ and $\underline{v}^1_{j,f,R}$ are pairwise independent and uniformly distributed on $\mathbb{F}_2^N$. Lemma 2 also tells us that $\underline{v}^2_{j_R,f,R}$ and $\underline{v}^2_{j,f,R}$ are equal and uniformly distributed on $\mathbb{F}_2^N$.

Define $\underline{v}^2_R$ as the common value taken by $\underline{v}^2_{j,f,R} = \underline{v}^2_{j_R,f,R}$. Joint AEP provides an asymptotically tight upper bound to each $P(E_j | j_R, j \in \mathcal{J}_{\{1\}})$ if we can show that

$$P(\mathbf{X}_{j,f,R}, \underline{Y}_R | j_R, j \in \mathcal{J}_{\{1\}}) = P(\mathbf{X}_{j,f,R} | j_R, j \in \mathcal{J}_{\{1\}}) P(\underline{Y}_R | j_R, j \in \mathcal{J}_{\{1\}}). \tag{45}$$

This is equivalent to showing that

$$P(\mathbf{X}_{j,f,R}, \underline{Y}_R | \underline{v}^2_R, j \in \mathcal{J}_{\{1\}}) = P(\mathbf{X}_{j,f,R} | \underline{v}^2_R, j \in \mathcal{J}_{\{1\}}) P(\underline{Y}_R | \underline{v}^2_R, j \in \mathcal{J}_{\{1\}}). \tag{46}$$

for each value of $\underline{v}^2_R$. Therefore, consider some arbitrary fixed $\underline{v}^2_R$. We can use (28) and the definition of conditional probability to get

$$P(\mathbf{X}_{j,f,R}, \underline{Y}_R | \underline{v}^2_R, j \in \mathcal{J}_{\{1\}})$$
$$= P(\mathbf{X}_{j,f,R} | \underline{v}^2_R, j \in \mathcal{J}_{\{1\}}) P(\underline{Y}_R | \mathbf{X}_{j,f,R}, \underline{v}^2_R, j \in \mathcal{J}_{\{1\}})$$
$$= P(\mathbf{X}_{j,f,R} | \underline{v}^2_R, j \in \mathcal{J}_{\{1\}}) P(\underline{Q}_R + \underline{W}_R | \underline{v}^1_{j,f,R}, \underline{v}^2_{j,f,R}, \underline{v}^2_R, j \in \mathcal{J}_{\{1\}}). \tag{47}$$

We see that, conditioned on $\underline{v}^2_R$, $\underline{q}_R$ is a random function of $\underline{v}^1_{j_R,f,R}$,

$$\underline{q}_R = g(\underline{v}^1_{j_R,f,R}; \underline{v}^2_R)$$

which is defined elementwise by Lemma 4. By Lemma 3, $\underline{v}^1_{j,f,R}$ and $\underline{v}^1_{j_R,f,R}$ are pairwise independent, therefore $\underline{v}^1_{j,f,R}$ and $\underline{q}_R$ are independent. Since $\underline{v}^2_{j,f,R} = \underline{v}^2_R$ we have

$$P(Event | \underline{v}^2_{j,f,R}, \underline{v}^2_R, j \in \mathcal{J}_{\{1\}}) = P(Event | \underline{v}^2_R, j \in \mathcal{J}_{\{1\}})$$

for any event. We conclude that

$$P(\underline{Q}_R + \underline{W}_R | \underline{v}^1_{j,f,R}, \underline{v}^2_{j,f,R}, \underline{v}^2_R, j \in \mathcal{J}_{\{1\}})$$
$$= P(\underline{Q}_R + \underline{W}_R | \underline{v}^2_{j,f,R}, \underline{v}^2_R, j \in \mathcal{J}_{\{1\}})$$
$$= P(\underline{Q}_R + \underline{W}_R | \underline{v}^2_R, j \in \mathcal{J}_{\{1\}}). \tag{48}$$

This allows us to conclude that (46) holds so we can use [18, Theorem 15.2.3] to get the following bound

$$P(E_j | j_R, j \in \mathcal{J}_{\{1\}}) < 2^{-N(I(Y_R; X^1_R | X^2_R) - 3\epsilon)}. \tag{49}$$

Similar steps can be used for the case when $j \in \mathcal{J}_{\{2\}}$ to get

$$P(E_j | j_R, j \in \mathcal{J}_{\{2\}}) < 2^{-N(I(Y_R; X^2_R | X^1_R) - 3\epsilon)}. \tag{50}$$

For the case when $j \in \mathcal{J}_{\{1,2\}}$, we have

$$\mathbf{X}_{j,f,R} = \mathbf{X}_{j_R,f,R} \oplus \begin{bmatrix} \underline{e}_{v,1} \\ \underline{e}_{v,1} \end{bmatrix}.$$

The most direct way to find a bound for this case is to reassign the address vectors so that this case is similar to the case when $j \in \mathcal{J}_{\{1\}}$. Define a binary matrix $\mathbf{D}_{\{1,2\}}$ given by

$$\mathbf{D}_{\{1,2\}} = \begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix}. \tag{51}$$

Then define a mapping function $\widetilde{\mathcal{M}}_{f,R}(\cdot)$ by

$$\widetilde{\mathcal{M}}_{f,R}(\underline{x}_{f,R}[n]) = \mathcal{M}_{f,R}(\mathbf{D}_{\{1,2\}}\underline{x}_{f,R}[n]). \tag{52}$$

Then define effective codeword matrices $\widetilde{\mathbf{X}}_{j_R,f,R}$ and $\widetilde{\mathbf{X}}_{j,f,R}$ by

$$\widetilde{\mathbf{X}}_{j_R,f,R} = \mathbf{D}_{\{1,2\}}\mathbf{X}_{j_R,f,R}$$

$$\widetilde{\mathbf{X}}_{j,f,R} = \mathbf{D}_{\{1,2\}}\mathbf{X}_{j,f,R} = \widetilde{\mathbf{X}}_{j_R,f,R} \oplus \begin{bmatrix} \underline{e}_{v,1} \\ \underline{0} \end{bmatrix}.$$

This is the same as the case where $j \in \mathcal{J}_{\{1\}}$ if the relay observes the $\underline{y}_R$ corresponding to codeword matricies, $\widetilde{\mathbf{X}}_{j_R,f,R}$ with the mapping function $\widetilde{\mathcal{M}}_{f,R}$. Therefore for the case where $j \in \mathcal{J}_{\{1,2\}}$, we have the bound

$$P(E_j|j_R, j \in \mathcal{J}_{\{1,2\}}) < 2^{-N(I(Y_R;\widetilde{X}_R^1|\widetilde{X}_R^2)-3\epsilon)} \tag{53}$$

which can be expressed in terms of the original address variables as

$$P(E_j|j_R, j \in \mathcal{J}_{\{1,2\}}) < 2^{-N(I(Y_R;X_R^1|X_R^1\oplus X_R^2)-3\epsilon)}. \tag{54}$$

By the definition of mutual information, we have

$$I(Y_R; X_R^1|X_R^1 \oplus X_R^2)$$

$$= H(Y_R|X_R^1 \oplus X_R^2) - H(Y_R|X_R^1, X_R^1 \oplus X_R^2)$$

$$= H(Y_R|X_R^1 \oplus X_R^2) - H(Y_R|X_R^1, X_R^2)$$

$$= I(Y_R; X_R^1, X_R^2|X_R^1 \oplus X_R^2).$$

Therefore the bound is equivalent to

$$P(E_j|j_R, j \in \mathcal{J}_{\{1,2\}}) < 2^{-N(I(Y_R;X_R^1,X_R^2|X_R^1\oplus X_R^2)-3\epsilon)}. \tag{55}$$

Lastly, for the case when $j \in \mathcal{J}_{\{1\}\{2\}}$, $\mathbf{X}_{j_R,f,R}$ and $\mathbf{X}_{j,f,R}$ are i.i.d. by Lemmas 2 and 3. We can therefore use joint AEP directly to get the bound

$$P(E_j|j_R, j \in \mathcal{J}_{\{1\}\{2\}}) < 2^{-N(I(Y_R;X_R^1,X_R^2)-3\epsilon)}. \tag{56}$$

Applying the upper bounds for each index set to (44), we get the following bound

$$P(Err|j_R) < \epsilon + \sum_{j \in \mathcal{J}_{\{1\}}} 2^{-N(I(Y_R; X_R^1 | X_R^2) - 3\epsilon)}$$

$$+ \sum_{j \in \mathcal{J}_{\{2\}}} 2^{-N(I(Y_R; X_R^2 | X_R^1) - 3\epsilon)}$$

$$+ \sum_{j \in \mathcal{J}_{\{1,2\}}} 2^{-N(I(Y_R; X_R^1, X_R^2 | X_R^1 \oplus X_R^2) - 3\epsilon)}$$

$$+ \sum_{j \in \mathcal{J}_{\{1\}\{2\}}} 2^{-N(I(Y_R; X_R^1, X_R^2) - 3\epsilon)}. \tag{57}$$

There are $2^{N\mathcal{R}} - 1$ elements in the sets $\mathcal{J}_{\{1\}}, \mathcal{J}_{\{2\}}$, and $\mathcal{J}_{\{1,2\}}$, and there are fewer than $2^{2N\mathcal{R}}$ elements in the last set $\mathcal{J}_{\{1\}\{2\}}$. Thus the upper bound on the probability of error for this code ensemble can be expressed

$$P(Err|j_R) < \epsilon + 2^{N(\mathcal{R} - I(Y_R; X_R^1 | X_R^2) + 3\epsilon)}$$

$$+ 2^{N(\mathcal{R} - I(Y_R; X_R^2 | X_R^1) + 3\epsilon)}$$

$$+ 2^{N(\mathcal{R} - I(Y_R; X_R^1, X_R^2 | X_R^1 \oplus X_R^2) + 3\epsilon)}$$

$$+ 2^{N(2\mathcal{R} - I(Y_R; X_R^1, X_R^2) + 3\epsilon)}. \tag{58}$$

Each of these terms can be made arbitrarily close to zero by increasing N as long as $\mathcal{R}$ satisfies

$$\mathcal{R} < \max(I(Y_R; X_R^1 | X_R^2), I(Y_R; X_R^2 | X_R^1), I(Y_R; X_R^1, X_R^2 | X_R^1 \oplus X_R^2), \frac{1}{2} I(Y_R; X_R^1, X_R^2)). \tag{59}$$

Note that this proof holds for an arbitrary j_R which means that the bound holds independent of the transmitted message.

Case $\{\ell \geq 2\}$: For a general ℓ, the proof is very similar. We split (39) into the disjoint classes of error events in (43) to get

$$P(Err|j_R) \leq P(\overline{E}_{j_R}|j_R) + \sum_{p=1}^{\ell} \sum_{\mathcal{S}_1, \ldots, \mathcal{S}_p} \sum_{j \in \mathcal{J}_{\mathcal{S}_1, \ldots, \mathcal{S}_p}} P(E_j|j_R, j \in \mathcal{J}_{\mathcal{S}_1, \ldots, \mathcal{S}_p}). \tag{60}$$

Then we find upper bounds on the probability of error for different classes of error events.

First, we consider the case where each $\mathcal{S}_m$, $m \in \{1, ..., p\}$ contains only its smallest element t_m. This first case is analogous to the case where $\{j \in \mathcal{J}_{\{1\}}\}$ for the proof when $\ell = 2$. By Lemmas 2 and 3, we have

$$P(\underline{V}_{j,f,R}^k = \underline{v}_1, \underline{V}_{j_R,f,R}^k = \underline{v}_2 | j \in \mathcal{J}_{\mathcal{S}_1, \ldots, \mathcal{S}_p}) = \begin{cases} 2^{-2N} & , k \in \mathcal{S} \\ 2^{-N} & , \underline{v}_1 = \underline{v}_2 \text{ and } k \notin \mathcal{S} \\ 0 & , \underline{v}_1 \neq \underline{v}_2 \text{ and } k \notin \mathcal{S}. \end{cases} \tag{61}$$

That is if $k \in \mathcal{S}$ then $\underline{v}_{j,f,R}^k$ and $\underline{v}_{j_R,f,R}^k$ are independent and uniformly distributed. If $k \notin \mathcal{S}$ they are equal and uniformly distributed. Let $\underline{v}_R^k$, $k \notin \mathcal{S}$ be the common value taken by the k_{th} row of $\mathbf{X}_{j,f,R}$ and $\mathbf{X}_{j_R,f,R}$.

The joint AEP gives an asymptotically tight upper bound to $P(E_j|j_R, j \in \mathcal{J}_{\mathcal{S}_1, \ldots, \mathcal{S}_p})$ if we can show that

$$P(\mathbf{X}_{j,f,R}, \underline{Y}_R | j_R, j \in \mathcal{J}_{\mathcal{S}_1, \ldots, \mathcal{S}_p}) = P(\mathbf{X}_{j,f,R} | j_R, j \in \mathcal{J}_{\mathcal{S}_1, \ldots, \mathcal{S}_p}) P(\underline{Y}_R | j_R, j \in \mathcal{J}_{\mathcal{S}_1, \ldots, \mathcal{S}_p}). \tag{62}$$

This is equivalent to showing that

$$P(\mathbf{X}_{j,f,R}, \underline{Y}_R | \{\underline{v}_R^k, k \notin \mathcal{S}\}, j \in \mathcal{J}_{\mathcal{S}_1,\dots,\mathcal{S}_p})$$
$$= P(\mathbf{X}_{j,f,R} | \{\underline{v}_R^k, k \notin \mathcal{S}\}, j \in \mathcal{J}_{\mathcal{S}_1,\dots,\mathcal{S}_p}) P(\underline{Y}_R | \{\underline{v}_R^k, k \notin \mathcal{S}\}, j \in \mathcal{J}_{\mathcal{S}_1,\dots,\mathcal{S}_p}) \tag{63}$$

for each possible set of values $\{\underline{v}_R^k, k \notin \mathcal{S}\}$. We can use (28) and the definition of conditional probability to get

$$P(\mathbf{X}_{j,f,R}, \underline{Y}_R | \{\underline{v}_R^k, k \notin \mathcal{S}\}, j \in \mathcal{J}_{\mathcal{S}_1,\dots,\mathcal{S}_p})$$
$$= P(\mathbf{X}_{j,f,R} | \{\underline{v}_R^k, k \notin \mathcal{S}\}, j \in \mathcal{J}_{\mathcal{S}_1,\dots,\mathcal{S}_p}) P(\underline{Y}_R | \mathbf{X}_{j,f,R}, \{\underline{v}_R^k, k \notin \mathcal{S}\}, j \in \mathcal{J}_{\mathcal{S}_1,\dots,\mathcal{S}_p})$$
$$= P(\mathbf{X}_{j,f,R} | \{\underline{v}_R^k, k \notin \mathcal{S}\}, j \in \mathcal{J}_{\mathcal{S}_1,\dots,\mathcal{S}_p}) P(\underline{Q}_R + \underline{W}_R | \mathbf{X}_{j,f,R}, \{\underline{v}_R^k, k \notin \mathcal{S}\}, j \in \mathcal{J}_{\mathcal{S}_1,\dots,\mathcal{S}_p}). \tag{64}$$

Thus the problem simplifies to showing that

$$P(\underline{Q}_R + \underline{W}_R | \mathbf{X}_{j,f,R}, \{\underline{v}_R^k, k \notin \mathcal{S}\}, j \in \mathcal{J}_{\mathcal{S}_1,\dots,\mathcal{S}_p})$$
$$= P(\underline{Q}_R + \underline{W}_R | \{\underline{v}_R^k, k \notin \mathcal{S}\}, j \in \mathcal{J}_{\mathcal{S}_1,\dots,\mathcal{S}_p}). \tag{65}$$

Since we are conditioning on $\{\{\underline{v}_R^k, k \notin \mathcal{S}\}, j \in \mathcal{J}_{\mathcal{S}_1,\dots,\mathcal{S}_p}\}$, the values taken by $\underline{v}_{j,f,R}^k, k \notin \mathcal{S}$ are already given. Therefore we have

$$P(\underline{Q}_R + \underline{W}_R | \mathbf{X}_{j,f,R}, \{\underline{v}_R^k, k \notin \mathcal{S}\}, j \in \mathcal{J}_{\mathcal{S}_1,\dots,\mathcal{S}_p})$$
$$= P(\underline{Q}_R + \underline{W}_R | \{\underline{v}_{j,f,R}^m, m \in \mathcal{S}\}, \{\underline{v}_R^k, k \notin \mathcal{S}\}, j \in \mathcal{J}_{\mathcal{S}_1,\dots,\mathcal{S}_p}).$$

The value taken by $\underline{q}_R$ conditioned on $\{\underline{v}_R^k, k \notin \mathcal{S}\}$ is a random function of $\{\underline{v}_{j_R,f,R}^m, m \in \mathcal{S}\}$,

$$\underline{q}_R = g(\{\underline{v}_{j_R,f,R}^m, m \in \mathcal{S}\}; \{\underline{v}_R^k, k \notin \mathcal{S}\})$$

which is defined element wise by Lemma 4. Therefore by the independence of $\{\underline{v}_{j,f,R}^m, m \in \mathcal{S}\}$ and $\{\underline{v}_{j_R,f,R}^m, m \in \mathcal{S}\}$ we can conclude that $\underline{q}_R$ is conditionally independent of $\{\underline{v}_{j,f,R}^m, m \in \mathcal{S}\}$ given $\{\{\underline{v}_R^k, k \notin \mathcal{S}\}, j \in \mathcal{J}_{\mathcal{S}_1,\dots,\mathcal{S}_p}\}$. Therefore since $\underline{w}_R$ is independent of any message, we can conclude that (65) holds. This is equivalent to (62) which allows us to apply joint AEP to get the upper bound

$$P(E_j | j_R, j \in \mathcal{J}_{\mathcal{S}_1,\dots,\mathcal{S}_p}) \leq 2^{-N(I(Y_R; \{X_R^m, m \in \mathcal{S}\} | \{X_R^k, k \notin \mathcal{S}\}) - 3\epsilon)}. \tag{66}$$

To extend this result to the case where each $\mathcal{S}_1, \dots, \mathcal{S}_p$ can contain multiple elements, we make this problem look like the first case. Define a matrix $\mathbf{D}_{\mathcal{S}_1,\dots,\mathcal{S}_p}$ whose m_{th} column $\underline{d}_m$ is given by

$$\underline{d}_{t_k}[n] = \begin{cases} 1, & n \in \mathcal{S}_k \\ 0, & n \notin \mathcal{S}_k \end{cases} \quad \forall k = 1, \dots, p$$

$$\underline{d}_m[n] = \begin{cases} 1, & n = m \\ 0, & n \neq m \end{cases} \quad \forall m \notin \mathcal{T}. \tag{67}$$

For example, if $\ell = 6$, $\mathcal{S}_1 = \{2, 4, 5\}$, and $\mathcal{S}_2 = \{3, 6\}$ we have

$$
\mathbf{D}_{\{2,4,5\},\{3,6\}} = \begin{bmatrix}
1 & 0 & 0 & 0 & 0 & 0 \\
0 & 1 & 0 & 0 & 0 & 0 \\
0 & 0 & 1 & 0 & 0 & 0 \\
0 & 1 & 0 & 1 & 0 & 0 \\
0 & 1 & 0 & 0 & 1 & 0 \\
0 & 0 & 1 & 0 & 0 & 1
\end{bmatrix}.
$$

Define effective codeword matrices $\widetilde{\mathbf{X}}_{j,f,R}$ and $\widetilde{\mathbf{X}}_{j_R,f,R}$ by

$$
\widetilde{\mathbf{X}}_{j,f,R} = \mathbf{D}_{\mathcal{S}_1,\ldots,\mathcal{S}_p} \mathbf{X}_{j,f,R}
$$

$$
\widetilde{\mathbf{X}}_{j_R,f,R} = \mathbf{D}_{\mathcal{S}_1,\ldots,\mathcal{S}_p} \mathbf{X}_{j_R,f,R}. \tag{68}
$$

Then the k_{th} row $\widetilde{\underline{v}}_{j,f,R}^k$ of $\widetilde{\mathbf{X}}_{j,f,R}$ is given by

$$
\widetilde{\underline{v}}_{j,f,R}^k = \begin{cases}
\widetilde{\underline{v}}_{j_R,f,R}^k & , \ k \notin \mathcal{T} \\
\widetilde{\underline{v}}_{j_R,f,R}^k \oplus \underline{e}_{v,m} & , \ k = t_m, \ m = 1, \ldots, p
\end{cases} \tag{69}
$$

for some set of pairwise independent error vectors $\underline{e}_{v,1}, \ldots, \underline{e}_{v,p} \in \mathbb{F}_2^N \setminus \{\underline{0}\}$.

For the $\ell = 6$ example, this means that

$$
\widetilde{\mathbf{X}}_{j,f,R} = \widetilde{\mathbf{X}}_{j_R,f,R} \oplus \begin{bmatrix}
\underline{0} \\
\underline{e}_{1,v} \\
\underline{e}_{2,v} \\
\underline{0} \\
\underline{0} \\
\underline{0}
\end{bmatrix}.
$$

This is the same as the case where each $\mathcal{S}_1, \ldots, \mathcal{S}_p$ contains only one element. Thus, we can apply the bound in (66) to get

$$
P(E_j | j_R, j \in \mathcal{J}_{\mathcal{S}_1,\ldots,\mathcal{S}_p}) \leq 2^{-N(I(Y_R;\{\widetilde{X}_R^k, \ k \in \mathcal{T}\}|\{\widetilde{X}_R^k, \ k \notin \mathcal{T}\}) - 3\epsilon)}. \tag{70}
$$

The only step that remains is to show that the mutual information in (70) can be expressed as

$$
I(Y_R; \{\widetilde{X}_R^k, \ k \in \mathcal{T}\}|\{\widetilde{X}_R^k, \ k \notin \mathcal{T}\})
$$
$$
= I(Y_R; \{X_R^k | k \in \mathcal{S}\}|\{X_R^k | k \notin \mathcal{S}\}, \{X_R^k \oplus Z_i | k \in \mathcal{S}_i\} \ \forall \ i = 1, \ldots, p) \tag{71}
$$

where $\mathcal{S} = \bigcup_{i=1}^p \mathcal{S}_i$, and each Z_i is an auxiliary random variable which is Bernoulli distributed with parameter $\frac{1}{2}$. By (68), we have

$$
\widetilde{x}_R^k = \begin{cases}
x_R^k & , \ k \in \overline{\mathcal{S}} \cup \mathcal{T} \\
x_R^k \oplus x_R^{t_m} & , \ k \in \mathcal{S}_m \setminus \{t_m\}.
\end{cases} \tag{72}
$$

We therefore have

$$\{\widetilde{x}_R^k | k \in \mathcal{T}\} \Leftrightarrow \{x_R^k | k \in \mathcal{T}\}$$

$$\{\widetilde{x}_R^k | k \notin \mathcal{T}\} \Leftrightarrow \{x_R^k | k \notin \mathcal{S}\} \cup \bigcup_{m=1}^{p} \{x_R^k \oplus x_R^{t_m} | k \in \mathcal{S}_m \setminus \{t_m\}\}.$$

The mutual information can therefore be expressed

$$I(Y_R; \{\widetilde{X}_R^k | k \in \mathcal{T}\} | \{\widetilde{X}_R^k | k \notin \mathcal{T}\})$$

$$= I(Y_R; \{X_R^k | k \in \mathcal{T}\} | \{X_R^k | k \notin \mathcal{S}\} \cup \bigcup_{m=1}^{p} \{X_R^k \oplus X_R^{t_m} | k \in \mathcal{S}_m \setminus \{t_m\}\})$$

$$= H(Y_R | \{X_R^k | k \notin \mathcal{S}\} \cup \bigcup_{m=1}^{p} \{X_R^k \oplus X_R^{t_m} | k \in \mathcal{S}_m \setminus \{t_m\}\}) - H(Y_R | X_R^1, ..., X_R^\ell).$$

The last equality follows because if we know $x_R^{t_m}$ and $x_R^k \oplus x_R^{t_m}$ then we know both $x_R^{t_m}$ and x_R^k. Which tells us that knowing $\{x_R^k | k \in \mathcal{T}\} \cup \bigcup_{m=1}^{p} \{x_R^k \oplus x_R^{t_m} | k \in \mathcal{S}_m \setminus \{t_m\}\}$ is equivalent to knowing $\{x_R^k | k \in \mathcal{S}\}$.

It can be shown for each $\mathcal{S}_m$, $m \in \{1, ..., p\}$ that

$$\{x_R^k \oplus x_R^{t_m} | k \in \mathcal{S}_m \setminus \{t_m\}\} \Leftrightarrow \{x_R^k \oplus z_m | k \in \mathcal{S}_m\}. \tag{73}$$

For example, if we consider our $\ell = 6$ case, we have $\mathcal{S}_1 = \{2, 4, 5\}$. If we know that

$$(x_R^4 \oplus x_R^2, x_R^5 \oplus x_R^2) = (a, b) \; , \; a, b \in \{0, 1\}$$

then we have

$$(x_R^2, x_R^4, x_R^5) \in \{(0, a, b), (1, \bar{a}, \bar{b})\}$$

which is equivalent to knowing

$$(x_R^2 \oplus z_1, x_R^4 \oplus z_1, x_R^5 \oplus z_1).$$

We therefore have

$$I(Y_R; \{\widetilde{X}_R^k | k \in \mathcal{T}\} | \{\widetilde{X}_R^k | k \notin \mathcal{T}\})$$

$$= H(Y_R | \{X_R^k | k \notin \mathcal{S}\} \cup \bigcup_{m=1}^{p} \{X_R^k \oplus X_R^{t_m} | k \in \mathcal{S}_m \setminus \{t_m\}\}) - H(Y_R | X_R^1, ..., X_R^\ell)$$

$$= H(Y_R | \{X_R^k | k \notin \mathcal{S}\} \cup \bigcup_{m=1}^{p} \{X_R^k \oplus Z_m | k \in \mathcal{S}_m\}) - H(Y_R | X_R^1, ..., X_R^\ell)$$

$$= I(Y_R; \{X_R^k | k \in \mathcal{S}\} | \{X_R^k | k \notin \mathcal{S}\}, \{X_R^k \oplus Z_i | k \in \mathcal{S}_i\} \; \forall \; i = 1, ..., p). \tag{74}$$

This is the same as (71), which allows us to restate the bound in (70) as

$$P(E_j | j_R, j \in \mathcal{J}_{\mathcal{S}_1, ..., \mathcal{S}_p})$$

$$\leq 2^{-N(I(Y_R; \{X_R^k | k \in \mathcal{S}\} | \{X_R^k | k \notin \mathcal{S}\}, \{X_R^k \oplus Z_i | k \in \mathcal{S}_i\} \; \forall \; i=1,...,p) - 3\epsilon)}$$

$$\triangleq 2^{-N(I(Y_R; \mathcal{S}_1, ..., \mathcal{S}_p) - 3\epsilon)}. \tag{75}$$

The last step defines a mutual information $I(Y_R; \mathcal{S}_1, ..., \mathcal{S}_p)$. This slight abuse of notation simplifies the last few steps of the proof.

Plugging this into (60), we have

$$P(Err|j_R) \leq P(\overline{E}_{j_R}|j_R) + \sum_{p=1}^{\ell} \sum_{\mathcal{S}_1, ..., \mathcal{S}_p} \sum_{j \in \mathcal{J}_{\mathcal{S}_1, ..., \mathcal{S}_p}} 2^{-N(I(Y_R; \mathcal{S}_1, ..., \mathcal{S}_p) - 3\epsilon)}. \tag{76}$$

For each possible $\mathcal{S}_1, ..., \mathcal{S}_p$ we have

$$|\mathcal{J}_{\mathcal{S}_1, ..., \mathcal{S}_p}| = (2^{N\mathcal{R}} - 1)(2^{N\mathcal{R}} - 2)...(2^{N\mathcal{R}} - p) < 2^{N\mathcal{R}p}.$$

Therefore we have

$$P(Err|j_R) \leq \epsilon + \sum_{p=1}^{\ell} \sum_{\mathcal{S}_1, ..., \mathcal{S}_p} 2^{N\mathcal{R}p} 2^{-N(I(Y_R; \mathcal{S}_1, ..., \mathcal{S}_p) - 3\epsilon)}$$

$$\leq \epsilon + \sum_{p=1}^{\ell} \sum_{\mathcal{S}_1, ..., \mathcal{S}_p} 2^{N(\mathcal{R}p - I(Y_R; \mathcal{S}_1, ..., \mathcal{S}_p) + 3\epsilon)}.$$

This bound approaches zero as long as

$$\mathcal{R} < \min_{\mathcal{S}, \overline{\mathcal{S}}, \mathcal{S}_1, ..., \mathcal{S}_p} \frac{1}{p} I(Y_R; \{X_R^k | k \in \mathcal{S}\} | \{X_R^k | k \in \overline{\mathcal{S}}\}, \{X_R^k \oplus Z_i | k \in \mathcal{S}_i\} \ \forall \ i \in \{1, ..., p\}). \tag{77}$$

This completes the proof.

References

[1] P. Popovski and T. Koike-Akino, "Coded bidirectional relaying in wireless networks," *New Directions in Wireless Communications Research*, pp. 291–316, 2009.

[2] M. P. Wilson, K. R. Narayanan, H. D. Pfister, and A. Sprintson, "Joint physical layer coding and network coding for bi-directional relaying," *IEEE Tran. Info. Theory*, vol. 56, pp. 5641–5654, Nov. 2010.

[3] B. Nazer and M. Gastpar, "Lattice coding increases multicast rates for Gaussian multiple-access networks," in *45th Annual Allerton Conference*, 2007.

[4] B. Nazer and M. Gastpar, "Reliable physical layer network coding," *Proceedings of the IEEE*, no. 99, pp. 1–23, 2011.

[5] S. Liew, S. Zhang, and L. Lu, "Physical-layer network coding: Tutorial, survey, and beyond," *Arxiv preprint arXiv:1105.4261*, 2011.

[6] W. Nam, S. Chung, and Y. Lee, "Capacity bounds for two-way relay channels," in *IEEE International Zurich Seminar on Communications*, pp. 144–147, IEEE, 2008.

[7] U. Erez and R. Zamir, "Achieving 1/2 log (1+ snr) on the awgn channel with lattice encoding and decoding," *IEEE Transactions on Information Theory*, vol. 50, no. 10, pp. 2293–2314, 2004.

[8] M. Wilson and K. Narayanan, "Power allocation strategies and lattice based coding schemes for bi-directional relaying," in *IEEE International Symposium on Information Theory*, pp. 344–348, IEEE, 2009.

[9] Y. Song and N. Devroye, "List decoding for nested lattices and applications to relay channels," *Arxiv preprint arXiv:1010.0182*, 2010.

[10] J. Zhan, U. Erez, M. Gastpar, and B. Nazer, "Mimo compute-and-forward," in *IEEE International Symposium on Information Theory*, pp. 2848–2852, IEEE, 2009.

[11] T. Koike-Akino, P. Popovski, and V. Tarokh, "Optimized constellations for two–way wireless relaying with physical network coding," *IEEE Journal on Selected Areas in Communications*, vol. 27, no. 5, pp. 773–787, 2009.

[12] B. Nazer and M. Gastpar, "Compute-and-forward: Harnessing interference through structured codes," *Arxiv preprint arXiv:0908.2119*, 2009.

[13] C. Feng, D. Silva, and F. Kschischang, "An algebraic approach to physical-layer network coding," *Arxiv preprint arXiv:1005.2646*, 2010.

[14] U. Wachsmann, R. Fischer, and J. Huber, "Multilevel codes: Theoretical concepts and practical design rules," *IEEE Transactions on Information Theory*, vol. 45, no. 5, pp. 1361–1391, 1999.

[15] R. Gallager, *Information theory and reliable communication*. John Wiley & Sons, Inc. New York, NY, USA, 1968.

[16] T. Richardson and R. Urbanke, *Modern coding theory*. Cambridge Univ Pr, 2008.

[17] A. Yedla, H. Pfister, and K. Narayanan, "Universality for the noisy slepian-wolf problem via spatial coupling," *Arxiv preprint arXiv:1105.6374*, 2011.

[18] T. Cover and J. Thomas, *Elements of information theory*. John Wiley and sons, 2006.

[19] R. Gallager, "A perspective on multiaccess channels," *Information Theory, IEEE Transactions on*, vol. 31, no. 2, pp. 124–142, 1985.